

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA**

Vânia Gomes de Almeida

**MODELAGEM E IDENTIFICAÇÃO AUTOMÁTICA DE CONSTRUÇÕES
DE ESTRUTURA ARGUMENTAL: UMA PROPOSTA PARA O CONSTRUCTICON
DA FRAMENET BRASIL**

Juiz de Fora
2022

Vânia Gomes de Almeida

**MODELAGEM E IDENTIFICAÇÃO AUTOMÁTICA DE CONSTRUÇÕES
DE ESTRUTURA ARGUMENTAL: UMA PROPOSTA PARA O CONSTRUCTICON
DA FRAMENET BRASIL**

Tese apresentada ao Programa de Pós-graduação em Linguística da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Linguística. Área de concentração: Linguística

Orientador: Dr. Tiago Timponi Torrent

Juiz de Fora
2022

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Almeida, Vânia Gomes de.

MODELAGEM E IDENTIFICAÇÃO AUTOMÁTICA DE CONSTRUÇÕES DE ESTRUTURA ARGUMENTAL: UMA PROPOSTA PARA O CONSTRUCTICON DA FRAMENET BRASIL / Vânia Gomes de Almeida. -- 2022.

216 p. : il.

Orientador: Tiago Timponi Torrent

Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Letras. Programa de Pós-Graduação em Linguística, 2022.

1. Modelagem de construções. 2. Reconhecimento de construções. 3. Estrutura argumental. 4. Gramática de Construções. 5. Constructicon. I. Torrent, Tiago Timponi, orient. II. Título.

Vânia Gomes de Almeida

**MODELAGEM E IDENTIFICAÇÃO AUTOMÁTICA DE CONSTRUÇÕES
DE ESTRUTURA ARGUMENTAL: UMA PROPOSTA PARA O CONSTRUCTICON
DA FRAMENET BRASIL**

Tese apresentada ao Programa de Pós
Graduação em Linguística da Universidade
Federal de Juiz de Fora como requisito parcial
à obtenção do título de Doutor em Linguística.
Área de concentração: Linguística.

Aprovada em 29 de julho de 2022

BANCA EXAMINADORA

Prof. Dr. Tiago Timponi Torrent - Orientador

Universidade Federal de Juiz de Fora

Profa. Dra. Luana Lopes Amaral

Universidade Federal de Minas Gerais

Prof. Dr. Sandro José Rigo

Universidade do Vale do Rio dos Sinos

Profa. Dra. Sandra Aparecida Faria de Almeida

Universidade Federal de Juiz de Fora

Prof. Dr. Ely Edison da Silva Matos

Universidade Federal de Juiz de Fora

DISCENTE

Vânia Gomes de Almeida

Juiz de Fora, 12 / 07 / 2022.



Documento assinado eletronicamente por **Sandro José rigo, Usuário Externo**, em 29/07/2022, às 11:48, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Tiago Timponi Torrent, Coordenador(a)**, em 29/07/2022, às 15:44, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sandra Aparecida Faria de Almeida, Professor(a)**, em 03/08/2022, às 10:34, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ely Edison da Silva Matos, Técnico Administrativo em Educação**, em 09/08/2022, às 10:58, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luana Lopes Amaral, Usuário Externo**, em 09/08/2022, às 11:21, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Vania Gomes de Almeida, Usuário Externo**, em 11/08/2022, às 15:43, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **0865849** e o código CRC **8040DE95**.

Às mulheres pretas e pobres deste país que a cada dia têm que encontrar novos motivos para seguir.

AGRADECIMENTOS

A Deus, por ser minha principal força motora, minha fortaleza e meu descanso. Obrigada por estar sempre comigo, cuidando, amando e consolando.

Agradeço à minha mãe Rosalice, por me mostrar que o amor é muito mais que palavras, amor é cuidado e consideração. Obrigada por me amar, mãe! Obrigada por me ensinar valores como responsabilidade e compromisso os quais eu tento carregar sempre comigo.

Ao meu pai Expedito, que quando me alfabetizou aos 4 anos não imaginava que sua filha ia se apaixonar pelas letras. Obrigada pai, por sempre me incentivar e me fazer acreditar que eu posso tudo.

À minha irmã Vanessa, por ser a melhor amiga que alguém poderia ter! Obrigada por ser companheira e meu exemplo de persistência!

Aos meus irmãos Luís Gustavo e Luiz Eduardo por compartilharem a vida comigo. À minha cunhada Tatiane pelas orações e aos meus sobrinhos Alan, Andressa, Maria Eduarda e Meirielen por serem motivos do meu mais sincero sorriso.

Um agradecimento especial ao meu orientador Tiago Timponi Torrent, não apenas por me orientar durante o mestrado e agora o doutorado, mas por durante esses 7 anos me ensinar sobre dedicação ao trabalho e a importância da pesquisa acadêmica. Obrigada por incentivar sempre o meu progresso e me motivar a buscar sempre mais. Obrigada por me incentivar e direcionar minha jornada acadêmica.

Aos meus amigos que estiveram ao meu lado em vários momentos de insegurança e incertezas. Ao Alexandre, por estar comigo me amparando e me apoiando. Obrigada Alexandre, por dividir comigo um ano de doutorado sanduíche. Obrigada apenas por estar presente, foi incrível viver essa experiência com você. À Malvina e Eulália, amigas que a graduação me trouxe, obrigada por estarem sempre perto quando eu mais preciso.

Agradeço ao grande time da FrameNet Brasil por me permitir vivenciar grandes experiências ao lado de vocês. Obrigada por me ensinarem tanto! Em especial agradeço ao Arthur pela grande contribuição com os experimentos computacionais.

À professora Patrícia Fabiane Amaral da Cunha Lacerda principal incentivadora por eu ter seguido os caminhos da pesquisa em linguística. Se agora um grande ciclo está se encerrando você tem grande participação nele.

Aos professores Luana, Sandro, Sandra e Ely por aceitarem fazer parte da banca examinadora desse trabalho. Aos professores Karen, Paulo Henrique, Patrícia e Regina por aceitarem compor a banca como suplente.

Agradeço à Universidade Federal de Juiz de Fora por ser realmente minha segunda casa durante esses 12 anos entre graduação e doutorado. Ao Programa de Pós-graduação em Linguística e à Faculdade de Letras por serem primordiais para o meu enriquecimento acadêmico e humano.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por contribuir significativamente para minha formação, bem como ao programa PROBRAL pela bolsa de doutorado sanduíche nas Universidades Leipzig e Dusseldorf, na Alemanha, (processo número 88887.185928/2018-00) que me permitiu, durante um ano, realizar trocas com outros pesquisadores, o que foi fundamental para o desenvolvimento desta pesquisa. Agradeço, portanto, aos meus orientadores na Alemanha, professor Alexander Ziem e Oliver Czulo, pelo suporte e dedicação.

“Eu só não queria significar
Porque significar limita a imaginação”
(BARROS, 2010)

RESUMO

Este trabalho tem por objetivo, em primeiro lugar, fornecer a modelagem linguístico computacional, no *Constructicon* da FrameNet Brasil, de 24 construções de estrutura argumental do Português Brasileiro, a saber: Sujeito-Predicado, Transitiva Direta Ativa, Transitiva Indireta, Transitiva Oblíqua, Bitransitiva, Intransitiva, Ergativa, Resultativa, Objeto Interdito, Argumento Cindido Artefato, Argumento Cindido Parte do Corpo, Argumento Cindido Posse Alienável, Argumento Cindido Entidade Atributo, Predicativa Nominal Atributiva, Predicativa Nominal Estativa, Predicativa Locativa, Verbo Nominal Objetiva, Mudança de Estado, Sem Sujeito Intransitiva, Existencial Apresentacional e Tempo Decorrido. Em segundo lugar, apresenta e discute os resultados de uma tarefa de identificação automática de construções cujo objetivo é o de identificar instâncias das construções acima. Nesse sentido, a proposta se enquadra nas discussões teórico-metodológicas da FrameNet Brasil (FN-Br) (SALOMÃO, 2009a), a qual desenvolve dois grandes recursos computacionais: um Lexicon e um Constructicon. O Lexicon tem por objetivos: criar uma representação computacional de frames, definidos por seus participantes e instrumentos e conectados entre si via relações entre frames; definir Unidades Lexicais, pareamentos entre um lema e um significado definido em termos de um frame; anotar sentenças que exemplifiquem os padrões de valência sintáticos e semânticos em que as ULs ocorrem. Já o Constructicon consiste em um recurso computacional no qual as informações estruturais sobre construções de uma língua são modeladas e é construído através de um modelo semântico baseado em frames, ligado às suas manifestações linguísticas (SALOMÃO, 2009a). Ambos os recursos se encontram interligados, na medida em que tanto o Lexicon como o Constructicon foram desenhados para lidar com todos os traços linguísticos relevantes para a identificação de um frame e de uma construção. Diante da interligação entre as duas frentes da FN-Br, apresentaremos uma proposta que pretende ampliar o escopo de pesquisa anterior (ALMEIDA, 2016), para auxiliar em tarefas de Compreensão de Língua Natural. Através da interação entre a Linguística e a Computação, este trabalho engloba as três fases apresentadas por Dias da Silva (2006) como necessárias para o desenvolvimento de um sistema em Linguística Computacional: a fase linguística, a fase linguístico computacional e a fase computacional. Para tanto, apresenta a especificação das construções na fase linguística, a representação dessas construções no *Constructicon*, na fase linguístico computacional, e propõe um sistema computacional para que essas construções possam ser reconhecidas na fase computacional. Tal sistema visa a testar a hipótese de que a modelagem de construções composta por informações relativas à sua forma e ao seu sentido, auxilia na

tarefa de reconhecimento de construtos por elas licenciados. Para tanto, instâncias das construções foram anotadas em *corpus* e submetidas a distintos modelos de Inteligência Artificial para serem classificadas conforme as construções existentes no *Constructicon*. Os resultados indicaram que, para construções muito frequentes no *corpus* de treinamento, a adição de informações advindas da modelagem afeta negativamente o desempenho do modelo. Por outro lado, para padrões construcionais menos frequentes, tal modelagem afeta positivamente o desempenho do modelo.

Palavras-chave: Modelagem de construções, Reconhecimento de construções, Estrutura argumental, Gramática de Construções, Constructicon.

ABSTRACT

This work aims, firstly, to provide the linguistic-computational modeling, in the *Constructicon* of FrameNet Brazil, of 24 Brazilian Portuguese argument structure constructions, namely: Subject Predicate, Active Direct Transitive, Indirect Transitive, Oblique Transitive, Ditransitive, Intransitive, Ergative, Resultative, Object Interrupted, Split Object Artifact, Split Object Body Part, Split Object Alienable Possession, Split Object Entity Attribute, Attributive Nominal Predicative, Stative Nominal Predicative, Locative Predicative, Objective Nominal Verb, State Change, Without Subject Intransitive, Existential Presentation and Elapsed Time. Second, it presents and discusses the results of an automatic construction identification task whose objective is to identify instances of the above constructions. In this sense, the proposal fits into the theoretical-methodological discussions of FrameNet Brazil (FN-Br) (SALOMÃO, 2009a), which develops two major computational resources: a Lexicon and a Constructicon. Lexicon aims to: create a computational representation of frames, defined by its participants and instruments, and connected to each other via relationships between frames; define Lexical Units, pairings between a lemma and a meaning defined in terms of a frame; note sentences that exemplify the syntactic and semantic valence patterns in which ULs occur. Constructicon, on the other hand, consists of a computational resource in which structural information about constructions of a language is modeled and it is built through a semantic model based on frames, linked to its linguistic manifestations (SALOMÃO, 2009a). Both resources are interconnected, as both Lexicon and Constructicon were designed to deal with all the linguistic features relevant to the identification of a frame and a construction. In view of the interconnection between the two fronts of the FN-Br, we will present a proposal that intends to expand the scope of previous research (ALMEIDA, 2016), to assist in Natural Language Comprehension tasks. Through the interaction between Linguistics and Computing, this work encompasses the three phases presented by Dias da Silva (2006) as necessary for the development of a system in Computational Linguistics: the linguistic phase, the linguistic-computational phase and the computational phase. Therefore, it presents the specification of constructions in the linguistic phase, the representation of these constructions in the *Constructicon*, in the linguistic-computational phase, and proposes a computational system so that these constructions can be recognized in the computational phase. This system aims to test the hypothesis that the modeling of constructions, composed of information related to their form and meaning, helps in the task of recognizing constructs licensed by them. For that, instances of the constructions were annotated in *corpus* and submitted to different models of Artificial

Intelligence to be classified according to the existing constructions in *Constructicon*. The results indicated that, for very frequent constructions in the training *corpus*, the addition of information from the modeling negatively affects the performance of the model. On the other hand, for less frequent constructional patterns, such modeling positively affects the performance of the model.

Keywords: Construction Modeling. Construction Recognition. Argument Structure. Construction Grammar. Constructicon.

LISTA DE ILUSTRAÇÕES

Figura 1 – Construção Bitransitiva	30
Figura 2 – Construção de Movimento Causado	31
Figura 3 – Construção de Mudança de Estado	32
Figura 4 – Esquema-x de pular.....	37
Figura 5 – Esquemas para a palavra <i>gato</i>	38
Figura 6 – Esquema Construção Movimento Direto	39
Figura 7 – Esquema Construção Movimento Direto	39
Figura 8 – Visão geral das estruturas e processos envolvidos na compreensão de linguagem baseada em simulação	40
Figura 9 – Possível estrutura transitória para o enunciado <i>he bakes</i>	44
Figura 10 – Construção Bitransitiva na FCG	46
Figura 11 – Construção DetHead	49
Figura 12 – Construto “O homem amanheceu doente” licenciado pela Construção de Mudança de Estado	51
Figura 13 – Construção HC.....	55
Figura 14 – Construção VP	55
Figura 15 – Construção VP representação alternativa.....	56
Figura 16 – <i>Frame</i> Quebrar.....	58
Figura 17 – Anotação lexicográfica do construto “O cabo da panela quebrou”	58
Figura 18 – Construção Ergativa	60
Figura 19 – Construção Sujeito_Predicado	61
Figura 20 – Construção Transitiva Direta Ativa	61
Figura 21 – Relação Evocação frame Ação_Transitiva.....	62
Figura 22 – Especificação no <i>Analyzer</i> da sentença <i>She gave him a cookie</i>	70
Figura 23 – Esquema de <i>ArrestedVBD</i>	73
Figura 24 – Construção <i>Active TransitiveVP</i>	74
Figura 25 – Construção <i>ExtractedNonFinitePassiveVP</i>	74
Figura 26 – Processamento <i>The cop arrested by detective was</i>	77
Figura 27 – Processamento de <i>The crook arrested by detective was</i>	78
Figura 28 – Diagrama com a ordem de aplicação das construções	80
Figura 29 – Método de Ativação Propagada	84
Figura 30 – Processamento do CARMA.....	85
Figura 31 – Exemplos de Construtos	86
Figura 32 – Análise do construto "O celular quebrou a tela." pelo CARMA	87
Figura 33 – Anotações SUC 2.0	91
Figura 34 – Exemplos de <i>n-grams</i> híbridos de SUC 2.0.....	91
Figura 35 – As instâncias de <i>vara_{VB} ute_{AB} och_{KN} VB</i>	92
Figura 36 – Exemplos de sentenças contendo a construção <i>Personal Pronoun + didn't + V + how</i>	99
Figura 37 – Exemplo do tratamento linguístico-computacional de uma configuração sintática	106
Figura 38 – Anotação no <i>UD_Portuguese-Bosque</i>	111
Figura 39 – Anotação construcional no <i>UD_Portuguese Bosque</i>	111
Figura 40 – Marcações feitas pelo modelo.....	115
Figura 41 – Fases de construção de um sistema em Processamento de Língua Natural	117
Figura 42 – Construção Sujeito_Predicado	121
Figura 43 – Construção Transitiva Direta Ativa	122

Figura 44 – Restrições aplicadas aos ECs da Construção Transitiva Direta Ativa	124
Figura 45 – Frame evocado pela Construção Transitiva.....	124
Figura 46 – Construção Transitiva Indireta.....	125
Figura 47 – Restrições aplicadas à Construção Transitiva Indireta.....	126
Figura 48 – Frame Transferir.....	126
Figura 49 – Relação evoca aplicada à Construção Transitiva Indireta.....	128
Figura 50 – Frame evocado pela Construção Transitiva Indireta.....	128
Figura 51 – Construção Transitiva Oblíqua.....	130
Figura 52 – Restrições aplicadas aos ECs da Construção Transitiva Oblíqua.....	131
Figura 53 – Construção Bitransitiva.....	133
Figura 54 – Frame evocado pela Construção Bitransitiva.....	134
Figura 55 – Restrições aplicadas aos ECs da Construção Bitransitiva.....	134
Figura 56 – Construção Intransitiva.....	135
Figura 57 – Restrições aplicadas aos ECs da Construção Intransitiva	135
Figura 58 – Frame evocado pela Construção Intransitiva.....	136
Figura 59 – Construção Ergativa	137
Figura 60 – Frame evocado pela Construção Ergativa	137
Figura 61 – Construção Objeto Interdito	140
Figura 62 – Restrições aplicadas aos ECs da Construção Objeto Interdito	140
Figura 63 – Frame evocado pela Construção Objeto Interdito.....	140
Figura 64 – CAC genérica.....	142
Figura 65 – Restrições aplicadas à CAC genérica.....	142
Figura 66 – Construção Argumento Cindido-artefato	143
Figura 67 – Restrições aplicadas aos ECs da CAC-artefato	143
Figura 68 – Relação Evoca CAC-parte do corpo	145
Figura 69 – Restrições aplicadas aos ECs da CAC-parte do corpo.....	145
Figura 70 – Construção Argumento Cindido-posse alienável.....	147
Figura 71 – Frame evocado pela CAC-entidade/atributo	148
Figura 72 – Construção Resultativa.....	155
Figura 73 – Restrições aplicadas aos ECs da Construção Resultativa.....	156
Figura 74 – Relação Evoca CAC-parte do corpo	157
Figura 75 – Construção Predicativa Nominal Atributiva.....	159
Figura 76 – Restrições aplicadas aos ECs da Construção Predicativa Nominal Atributiva ..	159
Figura 77 – Construção Predicativa Nominal Estativa	160
Figura 78 – Restrições aplicadas aos ECs da Construção Predicativa Nominal Estativa.....	160
Figura 79 – Construção Predicativa Locativa	161
Figura 80 – Restrições aplicadas aos ECs da Predicativa Locativa.....	162
Figura 81 – Construção Verbo-Nominal Objetiva.....	165
Figura 82 – Restrições aplicadas aos ECs da Verbo_Nominal_Objativa	165
Figura 83 – Frame evocado pela Construção Verbo_Nominal_Objativa	166
Figura 84 – Construção Mudança de Estado.....	168
Figura 85 – Restrições aplicadas aos ECs da Construção Mudança de Estado	169
Figura 86 – Frame evocado pela Construção Mudança de Estado.....	169
Figura 87 – Construções Herdeiras de Cláusula.....	171
Figura 88 – Construção Sem sujeito Intransitiva.....	172
Figura 89 – Frame evocado pela Construção Sem_Sujeito_Intransitiva	173
Figura 90 – Restrições aplicadas aos ECs da Construção Sem_Sujeito_Intransitiva.....	173
Figura 91 – Construção Existencial Apresentacional	174
Figura 92 – Restrições aplicadas aos ECs da Construção Existencial Apresentacional.....	175
Figura 93 – Frame evocado pela Construção Existencial Apresentacional	176

Figura 94 – Construção Tempo Decorrido.....	177
Figura 95 – Restrições aplicadas aos ECs da Construção Tempo Decorrido	177
Figura 96 – Relação Evoca Construção Tempo Decorrido.....	178
Figura 97 – Rede de construções de Estrutura Argumental	180

LISTA DE TABELAS

Tabela 1 – Modelos de sentenças analisadas e produzidas	82
Tabela 2 – Resultados obtidos pelo CARMA no reconhecimento de três construções de estrutura argumental.....	88
Tabela 3 – Padrões de candidatas a construção <i>geben</i>	96
Tabela 4 – Avaliação BERT	101
Tabela 5 – Resultados da <i>Edge Probing</i> no BERT Clone inferior e no CxGBERT inferior.	102
Tabela 6 – Avaliação do BERT Base	103
Tabela 7 – Número de anotações por construção	181
Tabela 8 – Resultados BERT-Base - Globais.....	183
Tabela 9 – Resultados BERT-Base Construções e ECs.....	183
Tabela 10 - Resultados BERT-Base + UD (rel & ordem) - Globais	185
Tabela 11 - Resultados BERT-Base + UD (rel & ordem) - Construções e ECs	185
Tabela 12 - Resultados BERT-Base + UD (rel & ordem) + Frame - Globais	186
Tabela 13 - Resultados BERT-Base + UD (rel & ordem) + Frame - Construções e ECs	186
Tabela 14 - Resultados Frames - Globais.....	187
Tabela 15 - Resultados Frames – Construções e Ecs.....	187
Tabela 16 -Resultados BERT-Base para as construções Transitiva Direta Ativa e Predicativa Nominal Atributiva	190
Tabela 17 - Resultados BERT-Base + UD (rel & ordem) para as construções Transitiva Direta Ativa e Predicativa Nominal Atributiva.....	191
Tabela 18 - BERT-Base + UD (rel & ordem) + Frame para as construções Transitiva Direta Ativa e Predicativa Nominal Atributiva.....	191
Tabela 19 - Resultados Frames para as construções Transitiva Direta Ativa e Predicativa Nominal Atributiva	192
Tabela 20 - Resultados para a construção Bitransitiva_ativa.....	193
Tabela 21 - Resultados para ECs da construção Bitransitiva_ativa.....	194
Tabela 22 - Resultados para a construção Ergativa	195
Tabela 23 - Resultados para os ECs da construção Ergativa	195
Tabela 24 - Resultados para a construção Intransitiva.....	197
Tabela 25 - Resultados para os ECs da construção Intransitiva	197
Tabela 26 - Resultados para a construção Predicativa_locativa.....	198
Tabela 27 - Resultados para os ECs da construção Predicativa_locativa	199
Tabela 28 - Resultados para a construção Predicativa_estativa	199
Tabela 29 - Resultados para os ECs da construção Predicativa_estativa	200
Tabela 30 - Resultados para a construção Transitiva_obliqua	201
Tabela 31 - Resultados para os ECs da construção Transitiva_obliqua	201

SUMÁRIO

1 INTRODUÇÃO.....	20
2 GRAMÁTICA DE CONSTRUÇÕES	25
2.1 FUNDAMENTOS DA GRAMÁTICA DE CONSTRUÇÕES	28
2.2 MODELOS BASEADOS EM UNIFICAÇÃO	33
2.2.1 Gramática de Construções Corporificada	34
2.2.2 Gramática de Construções Fluida.....	41
2.2.3 Gramática de Construções de Berkeley	47
2.3 REDES CONSTRUCIONAIS	53
2.4 IMPLEMENTAÇÕES COMPUTACIONAIS.....	56
2.4.1 O Lexicon	57
2.4.2 O Constructicon.....	59
3 LINGUÍSTICA COMPUTACIONAL E APREDIZAGEM DE MÁQUINA	65
3.1 COMPREENSÃO DE LÍNGUA NATURAL.....	66
3.2 TRABALHOS RELACIONADOS	69
3.2.1 ECG Analyzer.....	69
3.2.2 FCG Editor	78
3.2.3 CARMA	83
3.2.4 Swedish Constructicon	89
3.2.5 Mineração de construção.....	94
3.2.6 CxGBERT	97
4 METODOLOGIA	105
4.1 FASE LINGUÍSTICA	107
4.2 FASE LINGUÍSTICO-COMPUTACIONAL.....	108
4.1.1 Corpus Bosque da Universal Dependencies (UD) para o português.....	109
4.1.2. Anotação de instâncias do corpus no Constructicon	111
4.3 FASE COMPUTACIONAL	112
5 CONSTRUÇÕES DE ESTRUTURA ARGUMENTAL SUJEITO_PREDICADO ...	118
5.1 A CONSTRUÇÃO SUJEITO_PREDICADO	120
5.2 CONSTRUÇÃO TRANSITIVA DIRETA ATIVA	121
5.3 CONSTRUÇÃO TRANSITIVA INDIRETA	124
5.4 CONSTRUÇÃO TRANSITIVA OBLÍQUA.....	128
5.5 CONSTRUÇÃO BITRANSITIVA	131
5.6 CONSTRUÇÃO INTRANSITIVA.....	134
5.7 CONSTRUÇÃO ERGATIVA	136
5.8 CONSTRUÇÃO DE OBJETO INTERDITO.....	137
5.9 CONSTRUÇÃO DE ARGUMENTO CINDIDO.....	141
5.10 CONSTRUÇÃO RESULTATIVA.....	149
5.11 AS CONSTRUÇÕES PREDICATIVAS	157
5.12 COSNTRUÇÕES VERBOS-NOMINAIS.....	162
5.13 CONSTRUÇÕES CLAUSAIS SEM SUJEITO.....	169

6 RECONHECIMENTO E VALIDAÇÃO DE CONSTRUÇÕES	181
6.1 RESULTADOS E PRIMEIRA DISCUSSÃO.....	181
6.2 UMA NOVA ANÁLISE.....	193
6.3 CONCLUINDO A ANÁLISE.....	202
7 CONCLUSÃO	204
REFERÊNCIAS BIBLIOGRÁFICAS	209

1 INTRODUÇÃO

Esta pesquisa tem como objetivo a apresentação de uma proposta de investigação que foi desenvolvida durante o curso de Doutorado do Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora. Tal trabalho apresenta a modelagem de construções de estrutura argumental no Português Brasileiro e uma proposta para o reconhecimento automático dessas construções.

Esta proposta se enquadra nas discussões teórico-metodológicas da FrameNet Brasil (FN-Br) (SALOMÃO, 2009a). A FN-Br conta com dois grandes recursos computacionais: um *Lexicon* e um *Constructicon*. O *Lexicon* tem por objetivos: criar uma representação computacional de *frames*, definidos por seus participantes e instrumentos e conectados entre si via relações entre *frames*; definir Unidades Lexicais (ULs), pareamentos entre um lema e um significado definidos em termos de um *frame*; anotar sentenças que exemplifiquem os padrões de valência sintáticos e semânticos em que as ULs ocorrem. Já o *Constructicon* visa à criação de um recurso para a descrição das características semânticas e gramaticais de construções do PB, incorporando descrições interpretáveis computacionalmente para cada construção, oferecendo informações semânticas e especificando as relações entre as construções (FILLMORE ET AL., 2012). *Lexicon* e *Constructicon* se encontram conectados porque ambos descrevem satisfatoriamente as principais propriedades dos fenômenos linguísticos, na medida em que itens lexicais são construções e são, portanto, licenciados por construções lexicais, e que tanto itens lexicais quanto construções não lexicais podem evocar *frames*, que constituem o cerne da FrameNet

Diante da interligação entre as duas frentes da FN-Br, apresentamos uma proposta que pretende ampliar o que foi apresentado sobre o reconhecimento de construções por máquina em Almeida (2016), em que buscou-se apresentar a modelagem de construções do PB de modo a aprimorar o *Constructicon* da FN-Br e possibilitar, através dessa modelagem, o reconhecimento de construções por máquina. Em Almeida (2016), a modelagem realizada no *Constructicon* possibilitou o reconhecimento de construções pela máquina, agora nosso intuito é apresentar novas construções que ainda não haviam sido modeladas e apresentar uma tarefa de identificação dessas construções.

Através da interação entre a Linguística e a Computação, este trabalho engloba as três fases apresentadas por Dias da Silva (2006) como necessárias ao desenvolvimento de

um sistema em Linguística Computacional: a fase linguística, a fase linguístico computacional e a fase computacional.

No que tange à fase linguística, que concerne à especificação e à descrição das estruturas linguísticas conforme uma determinada teoria, nossos esforços concentram-se em identificar padrões construcionais já descritos, à luz da Gramática de Construções (FILLMORE, KAY & O'CONNOR, 1988; GOLDBERG, 1995; KAY & FILLMORE, 1999; FILLMORE, 2013). A fase linguístico-computacional foi realizada na medida em que essas estruturas identificadas na fase linguística recebam um tratamento computacional no *Constructicon* através da modelagem, e, no domínio da implementação, na fase computacional, pretendemos apresentar um sistema em que um algoritmo de Inteligência Artificial (IA), um classificador, conforme apresentado por Pustejovsky e Stubbs (2012), possa reconhecer as construções modeladas e anotadas em *corpus*.

Dessa maneira, conforme veremos ao longo do texto, os objetivos dessa pesquisa, além de enfocarem o tratamento linguístico dos padrões construcionais e descrevê-los através da formalização fornecida pelo *Lexicon* e pelo *Constructicon*, estarão voltados para a apresentação de um processo realizado através de um algoritmo em que as construções serão ou não identificadas. Essa proposta se justifica pela necessidade de ampliação do *Constructicon*, pelo fato de essa base apresentar dados estruturados com uma plausibilidade cognitiva, sendo assim, capaz de atuar em tarefas que vão além das fases linguísticas e representacionais.

Diante desse quadro, a nossa **hipótese é a de que, com a estrutura disponível na FN-Br, é possível que estruturas argumentais, que, segundo Goldberg (1995), correspondem a instâncias de construções que existem independentemente de verbos particulares e licenciam sentenças compostas por um verbo e seus argumentos, possam ser identificadas em *corpus* para fomentar tarefas em Compreensão de Língua Natural.**

Dado esse contexto, **o foco deste trabalho é modelar as construções de estrutura argumental do PB e apresentar uma proposta de identificação de tais construções, fundada na Semântica de *Frames* (FILLMORE, 1982; 1985) e na Gramática de Construções (FILLMORE, KAY & O'CONNOR, 1988; KAY & FILLMORE, 1999; FILLMORE, 2013), representada no *Constructicon* da FrameNet Brasil, e implementada em um algoritmo de IA.**

As implementações computacionais da FrameNet Brasil, o *Lexicon* e o *Constructicon* do Português Brasileiro, têm associado campos teóricos da Gramáticas das Construções e da Semântica de *Frames* a diversas tarefas de Linguística Computacional e Compreensão de Língua Natural (NLU), tais como construção de recursos lexicais e dicionários eletrônicos, identificação e catalogação de construções do PB e tradução por máquina (COSTA & TORRENT, 2017).

Dado esse contexto, a iniciativa de propor uma metodologia de modelagem e identificação de construções tira proveito da larga trajetória de desenvolvimento da FrameNet Brasil na busca por aliar o registro dos fenômenos linguísticos ao registro dos significados associados a esses fenômenos, no contexto de uso. Através de uma rede em que *frames*, elementos de *frame*, unidades lexicais, construções e elementos de construção representam o conhecimento linguístico, temos uma ferramenta flexível e abrangente para o estudo e a compreensão da língua sob o ponto de vista computacional que não deixa de lado aspectos centrais da cognição humana.

A intenção desta pesquisa surgiu, por um lado, da percepção de que o reconhecimento de construções por máquina, de acordo com suas propriedades sintático-semânticas, como foi iniciado em Almeida (2016), poderia ser ampliado. Por essa razão, este trabalho se justifica por uma proposição metodológica em que o tratamento linguístico-computacional dado a novos padrões construcionais possa levar ao seu reconhecimento em *corpus* a partir da modelagem realizada no *Constructicon*.

Por outro lado, a proposta de desenvolvimento de uma metodologia semiautomatizada para identificação de padrões construcionais, pode tirar proveito do grande número de dados disponíveis dentro dos diferentes campos da Linguística Computacional. Allen (1995) afirma que o trabalho em linguística computacional utiliza de várias ferramentas da Inteligência Artificial, como algoritmos, estrutura de dados modelos formais para representação do conhecimento e outros. Dentre esses modelos, há os métodos semisupervisionados de aprendizagem de máquina (PUSTEJOVSKY & STUBBS, 2012), nos quais, partindo de uma estrutura existente, algoritmos de IA podem reconhecer novas estruturas em potencial. Assim, em princípio, a partir de um *corpus* anotado para construções, novas construções poderiam ser identificadas.

A nossa proposta conta com um modelo cognitivamente plausível de uma rede construcional – o *Constructicon* – já que propõe que construções linguísticas, enquanto um pareamento de forma e sentido, podem ser tratadas sobre o aspecto computacional, uma vez que conseguimos demonstrar sintática e semanticamente as principais

características dessas estruturas, incluindo, nessa rede, os mais diversos tipos de construções linguísticas com o tratamento computacional fornecido pelo *Constructicon*.

Essa estratégia metodológica se apresenta como uma ferramenta importante que pode auxiliar em tarefas de Compreensão de Língua Natural. De um lado, temos modelos cognitivamente plausíveis como o *Constructicon*, capaz de fornecer informações a partir de um viés computacional para construções. Por outro lado, modelos semiautomatizados se apresentam como uma alternativa encorajadora, já que será possível associar a estrutura de dados da FN-Br, que possui modelagem e anotação realizada por humanos, juntamente com o reconhecimento automático realizado por algoritmos de IA atuando sobre um *corpus*.

Assim, o contexto de pesquisa em que este trabalho está alicerçado verifica em que medida a estrutura presente na FN-Br é suficiente para colaborar com tarefas de Compreensão de Língua Natural, e em que medida outro tipo de estrutura de dados se faz necessária.

Este trabalho, portanto, se propõe os seguintes objetivos adjacentes que se fazem relevantes, a saber:

- (i) Desenvolvimento de um repositório amplo de construções do PB;
- (ii) Desenvolvimento de um modelo de treinamento probabilístico semisupervisionado para identificação de padrões construcionais.

Os resultados nos apontam que para construções muito frequentes no *corpus* de treinamento, a adição de informações advindas da modelagem afeta negativamente o desempenho do modelo. Por outro lado, para padrões construcionais menos frequentes, tal modelagem afeta positivamente o desempenho do modelo.

A fim de situar a hipótese e o objetivo, neste trabalho, no capítulo 2, trazemos a principal teoria que alimenta essa pesquisa, a Gramática de Construções e suas diferentes abordagens e aplicações. Já no capítulo 3, delineamos a teoria da Linguística Computacional e apresentamos alguns trabalhos relacionados ao reconhecimento de construções, enquanto, no capítulo 4, apresentamos a metodologia seguida neste trabalho, indicando como as três fases apontadas por Dias da Silva (1996, 2006) são aplicadas, juntamente com anotação em *corpus* das construções modeladas. Posteriormente, no capítulo 5, temos a modelagem das construções de estrutura argumental, composta das primeiras análises feitas em Almeida (2016), acrescidas com novos modelos, e, por fim, no capítulo 6 os resultados de uma tarefa de identificação de construções são discutidos.

Expostos os conteúdos que serão abordados nesse trabalho, passemos agora à principal teoria que embasa este trabalho, apresentando suas diferentes abordagens e as aplicações computacionais dessa teoria

2 GRAMÁTICA DE CONSTRUÇÕES

A Linguística como ciência autônoma nasce com Saussure no início do século XX com a definição da língua como objeto de estudo. Desde então, os estudos linguísticos têm abarcado várias vertentes de pensamentos e linhas de pesquisas que tratam da linguagem em perspectivas diferentes, fornecendo à Linguística uma história permeada por grandes mudanças de paradigmas.

Ao propor uma série de dicotomias, Saussure apresenta uma visão de língua como fenômeno social e analisável a partir de um sistema de signos. Em seu Curso de Linguística Geral, desenvolvido entre 1911-1913 e publicado postumamente em 1916 por seus alunos, encontramos duas posições que são de suma importância para a compreensão dos estudos linguísticos e constituição do seu objeto em diferentes perspectivas, a saber: a natureza do objeto da linguística e a natureza do significante linguístico. A primeira diz respeito à delimitação e a concepção de uma teoria para o objeto linguístico. A segunda, considera que os significantes existem na língua, ou por uma relação de combinação ou convenção por partes dos falantes de uma língua (SAUSSURE, 1972 [1916]). Devido ao seu recorte sincrônico e sua visão sistemática, os estudos saussureanos priorizaram a análise formal pelo estudo disciplinar do significante, negligenciando a sociedade, o sujeito, a cognição e o funcionamento discursivo da língua.

Por volta da segunda metade do século XX, os trabalhos de Noam Chomsky desenvolvem a Linguística Gerativa, dando aos estudos formais da linguagem refinamento analítico e epistemológico. Enquanto para Saussure a linguagem é uma instituição social e uma convenção social, para Chomsky a linguagem é uma faculdade mental inata e transmitida geneticamente pela espécie. O objeto de investigação passa a ser então os estados da mente/cérebro que fazem parte do comportamento linguístico.

Ao eleger a sintaxe como o cerne do processamento linguístico, a teoria gerativa de Noam Chomsky se destacou justamente por considerar a manipulação do sentido através do estudo da forma, ou seja, o significado é imanente. Nessas perspectivas, a informação semântica estava contida no léxico, mais especificamente na estrutura profunda, caracterizando a estruturas gramaticais por um conjunto de regras de integração semântica (CHOMSKY, 1965). Os estudos científicos dedicados à manipulação de formas se ocuparam do processamento de exemplos artificiais que tinham correlações fracas entre forma e significado, uma vez que esse era reduzido a condições de verdade e necessidade.

A semântica formal, assumindo uma abordagem objetivista do significado, ao propor que a atribuição da linguagem seria a de descrever o estado de coisas no mundo, acaba, portanto, por se fundar sobre um falante-ouvinte inocente (FILLMORE, 1979), que processa o significado das sentenças de acordo com as suas partes e a sua organização. O falante-ouvinte inocente, apesar de conhecer as regras sintáticas da língua, não tem conhecimento sobre expressões idiomáticas ou usos metafóricos, processando-as literal e composicionalmente.

Assim, na teoria chomskyana, o significado de uma sentença é o tipo de situação que ela descreve e a descrição dessas situações possíveis é equivalente às condições de verdade da sentença, ou seja, o significado é concebido como reflexo da realidade. O processamento semântico realizado pela semântica formal acontece independentemente do contexto.

Contudo, apesar da sofisticação teórica que recebem os estudos da sintaxe, o modelo gerativo passa a ser questionado, principalmente, devido à aceção de que humanos possuem habilidades inatas que os capacitam a aprender e usar uma ou mais línguas. A crítica recai principalmente sobre a proposta de que essas habilidades e estruturas são específicas da linguagem.

Apesar disso, ao mencionar que a teoria linguística é mentalista, Chomsky faz o primeiro rompimento com a tradição estruturalista em direção ao surgimento da Linguística Cognitiva, demonstrando de fato a importância, para a compreensão da linguagem, dos fenômenos da natureza cognitiva que estão relacionados com o modo como a mente interage com o mundo físico, bem como os processos mediados por essa interação. No entanto, os estudos ligados ao significado ainda ficavam restritos à forma lógica, não abrangendo traços próprios do uso concreto da língua pelos falantes.

Nas décadas de 60 e 70, muitos pesquisadores, como George Lakoff e Charles Fillmore, se posicionaram contra a tradição chomskyana, abrindo caminho para uma nova abordagem do fenômeno da linguagem. Nesse contexto, surgem, então, os primeiros trabalhos em Linguística Cognitiva (NEWMAYER, 1986).

Essa nova abordagem não considera a linguagem como componente autônomo da mente, mas defende uma visão integradora não fazendo distinção entre o conhecimento linguístico e não-linguístico, considerando que alguns processos cognitivos humanos básicos vão estar relacionados à utilização de estruturas linguísticas em contextos reais nos quais essas estruturas são construídas. Temos então, nessa visão teórica, uma relação entre pensamento, linguagem e experiência.

Além disso, enquanto, no gerativismo, a sintaxe assumia destaque nas análises, na Linguística Cognitiva, busca-se muito mais o trabalho com a significação, já que se focaliza a linguagem como uma forma de ação, ou seja, não há significados prontos, mas mecanismos de construção de sentidos a partir de dados contextuais ricos e dinâmicos (FILLMORE, 1985).

Nesse sentido, os autores que se opunham ao gerativismo buscavam uma perspectiva teórica que considerasse as relações entre sintaxe e semântica, investigando especialmente as relações entre forma e significado. De acordo com Salomão (2007), há um amplo escopo analítico disponível para o trabalho em Linguística Cognitiva, que culmina no desenvolvimento de teorias que, embora tenham uma base teórica comum, adotam objetos de pesquisa diferentes e assumem metodologias e perspectivas variadas.

Segundo Salomão (2007), algumas teorias que ganharam destaque em Linguística Cognitiva foram: o trabalho de George Lakoff sobre os processos linguísticos de Categorização (LAKOFF, 1987), a proposição da Semântica de Frames por Charles Fillmore (FILLMORE, 1977a, 1977b, 1982, 1985), a psicologia da Gestalt aplicada à Dinâmica de Forças de Talmy (1978,1983) e à Gramática Cognitiva de Langacker (1987, 1991), a Teoria Conceptual da Metáfora (LAKOFF, 1983; LAKOFF; JOHNSON, 1980, 1999, 2002), o trabalho de Gilles Fauconnier, Eve Sweetser e Mark Turner, sobre a teoria dos Espaços Mentais e o processo cognitivo de Mesclagem (FAUCONNIER, 1994, 1997; FAUCONNIER; SWEETSER, 1996; FAUCONNIER; TURNER, 2002; DANCYGIER; SWEETSER, 2005).

Um dos empreendimentos da Linguística Cognitiva é o tratamento de construções gramaticais, que levou ao desenvolvimento da Gramática de Construções (KAY e FILLMORE, 1999). Através da permissão de que o léxico e a sintaxe formam um contínuo de construções gramaticais, ancorado na proposição de um pareamento entre forma e significado, a Gramática de Construções assume que esses padrões de forma e significado compreendem características compartilhadas por instanciações específicas variadas. Desse modo, encontramos um modelo de gramática baseado em esquemas gerais abstratos, diferente de um modelo baseado na manipulação de formas através de regras, como os modelos gerativos.

Apesar da heterogeneidade presente na Linguística Cognitiva, com relação às formulações teóricas e às ferramentas analíticas adotadas, Salomão (2009) afirma que há três princípios fundamentais que são comuns a qualquer perspectiva dentro dessa linha

de pesquisa e que, conseqüentemente, fundamentam os estudos em Gramáticas de Construções (GC) são eles:

1. A cognição linguística é contínua aos demais sistemas cognitivos; portanto, a linguagem não é um sistema cognitivo autônomo.
2. A gramática é uma grande rede de construções; portanto, postula-se uma continuidade básica entre sintaxe e léxico, calcada no uso linguístico.
3. Todo processo de significação procede pela projeção entre domínios cognitivos; portanto, a semântica cognitivista tem um viés inferencialista, que a diferencia do referencialismo da ortodoxia. (SALOMÃO, 2009. p. 22)

Nas próximas seções, serão apresentados os fundamentos da Gramática de Construções e a distinção de modelos dentro dessa teoria.

2.1 FUNDAMENTOS DA GRAMÁTICA DE CONSTRUÇÕES

A Gramática de Construções é um desenvolvimento recente na teoria linguística, começou a ser discutida mais intensamente no final dos anos 80 e início dos anos 90 com os trabalhos de Fillmore, Kay e O'Connor (1988), Goldberg (1995), Kay e Fillmore (1999) e Croft (2001).

Nessa abordagem, a gramática de uma língua consiste em padrões para a constituição de expressões, os quais são referidos como construções. Sendo assim, as principais acepções da Gramática de Construções são: construções são unidades básicas da língua que se constituem em correspondências entre forma e significado (GOLDBERG, 1995; KAY e FILLMORE, 1999); qualquer material linguístico desde o mais simples ao mais complexo pode ser considerado uma construção, havendo continuidade entre o léxico e a sintaxe, abandonando a ideia de modularidade; a gramática de uma língua consiste de uma rede de construções mediada por diferentes relações (FILLMORE, 2008).

Um estudo inicial sobre construções partiu de Lakoff (1987) sobre as construções dêiticas com *there*, em que autor apresenta a existência de uma instância prototípica e de várias subconstruções relacionados ao núcleo *there*. Essa construção representa uma sentença com estrutura sintática THERE_X em que o dêitico de lugar *there* ocorre sempre em posição inicial.

O estudo de Lakoff (1987) mapeou uma série de construções, em que temos a existência de uma Construção Dêitica Central como em (1).

- (1) There's Harry with the red jacket on.

Essa construção indica que uma determinada entidade existe e está presente em lugar acessível ao campo visual do falante: em (1) o dêitico aponta para “*Harry with the red jacket on*”.

A partir da Construção Dêitica Central, outras construções podem ser identificadas como é o caso da Construção Existencial com *there* em (2).

- (2) There's been a man shot.

A diferença entre essas construções é que, enquanto a Construção Dêitica indica um local específico, a Construção Existencial não, havendo, assim, um contraste do dêitico *there* com o dêitico *here*, no primeiro caso, que não é observado na Construção Existencial.

Lakoff (1987) pontua que uma construção motiva a outra, afirmando que uma dada construção é motivada na medida em que sua estrutura é herdada de outras construções na língua. O objetivo de se propor uma relação de motivação se relaciona a uma tentativa de garantir à teoria adequação explicativa, uma vez que relações de motivação e/ou herança capturam generalidades observáveis nos fenômenos linguísticos.

Inspirada no trabalho de Lakoff, Adele Goldberg (1995) inaugura uma nova etapa dos estudos construcionistas ao analisar as chamadas Construções de Estrutura Argumental, que licenciam sentenças compostas por um verbo e seus argumentos. Através dessa análise, propõe-se uma distinção entre os papéis argumentais, que são associados às construções e os papéis participantes, associados aos verbos, ou seja, o verbo perfila os papéis participantes, já os papéis argumentais são as possibilidades previstas pela construção, mas não necessariamente pelo verbo.

Conforme foi mencionado acima, segundo Goldberg, construções são definidas como pareamentos convencionais e aprendidos de forma e função, em variados níveis de complexidade e abstração (GOLDBERG, 1995, 2006). Essa definição dada por Goldberg, segundo a própria linguista, intenciona destacar a semelhança entre palavras e unidades sintagmáticas mais amplas. Dessa forma, construções são caracterizadas como tais independentemente do tamanho de sua estrutura, podendo, assim, corresponderem a uma palavra, uma expressão idiomática ou mesmo a padrões linguísticos mais gerais. Isso

significa, portanto, que, na perspectiva de Goldberg da Gramática de Construções, assim como nas demais perspectivas construcionistas, não se assume uma divisão estrita entre construções sintáticas e lexicais, ou seja, essas construções podem diferir em complexidade interna, mas mantêm a relação entre forma e significado na sua constituição.

Para uma análise construcional, a autora propõe um modelo baseado na interação entre semântica verbal e semântica construcional. Com esse modelo, não é necessário apresentar vários sentidos verbais para explicar a diferença entre as construções.

Com base nessa integração, Goldberg (1995) focou suas análises em um conjunto de construções do inglês, entre as quais se destacam construções como a de Movimento Causado e a Bitransitiva, esta última exemplificada em (3) e esquematizada na Figura 1.

(3) Pedro deu flores a Maria

Figura 1 – Construção Bitransitiva

Sem:	CAUSAR RECEBER	<Agt >	< Tema >	< Recip. >
	↓	↓	↓	↓
R:	DAR	<Doador>	<Donativo>	<Destinatário>
	↓	↓	↓	↓
Syn	V	<Subj>	<ObjD>	<ObjIn>

Fonte: Traduzido e adaptado pela autora a partir de Goldberg (1995, p. 50)

A formalização de construções na abordagem de Goldberg é representada através de caixas, em que são mapeadas as características sintáticas e semânticas das construções. A Figura 1 representa a Construção Bitransitiva, em que, na primeira linha, temos a semântica da construção (Sem), com os papéis semânticos dos argumentos da construção; na segunda (R), o predicador, no caso, um verbo, que perfila os papéis participantes; e, na terceira linha, as funções gramaticais atribuídas à fusão de participantes e argumentos na expressão sintática (Syn).

A construção Bitransitiva, que licencia construtos como (3), apresenta uma semântica em que o predicador *dar* seleciona três argumentos, *Pedro, flores e Maria*, com

função gramatical de *sujeito*, *objeto direto* e *objeto indireto* respectivamente, e com o papel semântico de *agente*, *tema* e *recipiente*, configurados cognitivamente pela construção, e a especificação de *doador* e *donativo* fornecida pelo predicador. O papel do destinatário pode ser contribuído pela construção; isso é indicado pela linha tracejada entre a função de argumento do destinatário e a matriz de predicado funções dos participantes. Ou seja, pode haver verbos não-bitransitivos que ocorram nessa construção.

Outra construção analisada por Goldberg é a Construção de Movimento Causado, exemplificada em (4) e representada na Figura 2.

(4) Pedro levou Maria para o jardim.

Figura 2 – Construção de Movimento Causado

	Sem CAUSAR MOVER	<Causa >	< Tema >	< Alvo>
	↓	↓	↓	↓
R:	LEVAR	<Levador>	<Levado>	<Destino>
	↓	↓	↓	↓
Syn	V	<Subj>	<Obj>	<Obj >

Fonte: Traduzido e adaptado pela autora a partir de Goldberg (1995, p. 52)

No caso da Construção de Movimento Causado, que apresenta instanciações como (4), encontramos três argumentos, *Pedro*, *Maria* e *jardim*, com funções gramaticais de *sujeito*, *objeto direto* e *oblíquo* respectivamente. O papel semântico de *Pedro* é de *Causa*, da *Maria* de *Tema* e *jardim* de *Alvo* com papéis de *levador*, *levado* e *destino* especificados pelo predicador.

As principais vantagens desse modelo de análise se referem ao mapeamento de generalizações e a economicidade, uma vez que diferentes verbos podem estar relacionados a uma mesma construção, porque o significado está presente também na construção e não apenas no verbo. Vejam-se, por exemplo, os construtos (5), (6) e (7), licenciados pela Construção Mudança de Estado, que será apresentada no capítulo 5.

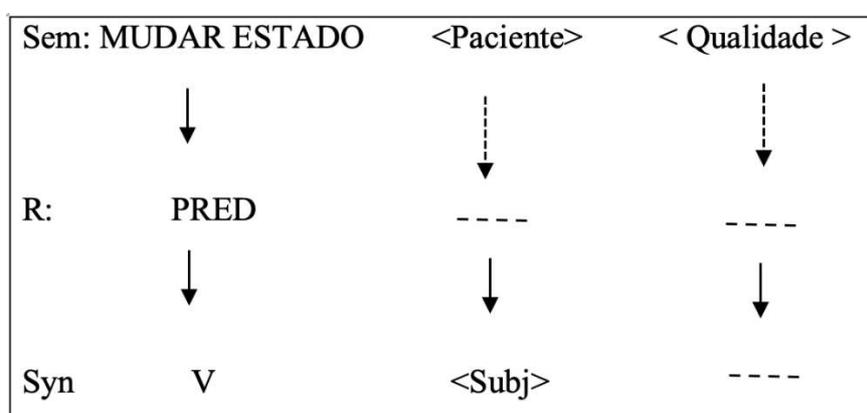
(5) Jack amanheceu morto.

- (6) João acabou condenado.
 (7) Amy apareceu bêbada.

A Construção de Mudança de Estado é um tipo de construção de estrutura argumental caracterizada por [SN [V [SAdj]]]. Como podemos perceber, essa construção apresenta uma predicação adjetival que é expressa pelo predicado verbo-nominal.

Ao propor uma representação dessa construção no modelo de Goldberg teríamos o esquema disposto na Figura 3.

Figura 3 – Construção de Mudança de Estado



Fonte: Criado pela autora

Como podemos observar, o verbo *amanhecer* é um verbo que caracteriza um fenômeno da natureza, sendo assim, não seleciona nenhum argumento, e os verbos *aparecer* e *acordar* selecionam apenas um argumento. Isso evidencia que, se a semântica fosse exclusivamente lexical, sentenças como (5), (6) e (7) requereriam que entradas lexicais adicionais fossem criadas para esses verbos. O que Goldberg (1995; 2010) defende é um sentido global da construção que resulta da associação entre o significado da construção e o significado dos itens lexicais que compõem a construção. Assim, na Figura 3, percebemos que os papéis semânticos de *Paciente* e de *Qualidade* são contribuições da construção de Mudança de Estado. A Construção Mudança de Estado será descrita mais explicitamente no Capítulo 5.

Por isso, a autora afirma que as construções têm um significado próprio independente dos verbos ou itens lexicais que as compõem, sendo que as palavras apenas contribuem, mas não são responsáveis por todo significado. No caso da construção acima,

os verbos fornecem informações sobre como se deu o processo de mudança de estado: se ao longo de uma noite, se ao final de um período ou se repentinamente.

O modelo de Goldberg (1995) para a Gramática das Construções desenvolveu-se em uma produtiva interface com a área de aquisição da linguagem (GOLDBERG, 2006). Esse modelo, entretanto, possui limitações de formalização, que tornam difícil sua aplicação para além da fase linguística do processo de modelagem computacional. Nesse sentido, o modelo da Gramática de Construções baseado em unificação proposto por Fillmore, Kay e O'Connor (1988), por Kay e Fillmore (1999), Fillmore (2008) e Fillmore (2013), a Gramática de Construções de Berkeley (BCG), revela-se mais adequado. Este será objeto da nossa próxima seção.

2.2 MODELOS BASEADOS EM UNIFICAÇÃO

Outros modelos de Gramática de Construções seguem paralelamente ao modelo de Goldberg, como a *Embodied Construction Grammar* – ECG, (DODGE, 2010; BERGEN & CHANG, 2013) a *Fluid Construction Grammar* – FCG (STEELS, 2013) e a *Berkeley Construction Grammar* – BCG (KAY & FILLMORE, 1999; FILLMORE, 2008; FILLMORE, 2013), modelos baseados em unificação que se fundamentam matematicamente e buscam incorporar práticas computacionais para tratar o significado das expressões linguísticas. Para Kay e Fillmore (1999, p.8), uma abordagem construcional assume o compromisso de que tanto os padrões relativamente gerais da língua como os padrões mais idiomáticos devam receber o mesmo tratamento. Assim, é necessário desenvolver um sistema de representação capaz de codificar economicamente e sem perda de generalização todas as construções (ou padrões) da língua, do mais idiomático ao mais geral.

O que os modelos citados têm em comum é a tentativa de modelar computacionalmente a cognição humana e os mecanismos que fundamentam o comportamento linguístico humano. Na busca por explicar o uso da linguagem humana, ou seja, os procedimentos de representação que os seres humanos usam para aprender, produzir e compreender os enunciados, essas abordagens têm no significado um papel crucial. Muitas vezes negligenciado em outros formalismos, o significado, aqui, deve ser incorporado às representações gramaticais com o fim de demonstrar como um modelo baseado em construções é cognitivamente plausível.

2.2.1 Gramática de Construções Corporificada

A *Embodied Construction Grammar* (ECG), Gramática de Construções Corporificada, compartilha com outras abordagens construcionistas o interesse em como construções de vários tipos contribuem no significado e na função dos enunciados, isto é, como elas são usadas.

Segundo Dodge (2010), a ECG foi projetada com os princípios da Teoria Neural da Linguagem (NTL), uma vez que essa teoria traz implicações importantes sobre quais aspectos do entendimento dos enunciados precisam ou não ser especificados. Dodge (2010) afirma que as gramáticas das línguas são constituídas de construções e esquemas em que construções são pareamento de forma e significado e as restrições de significado são definidos em termos de esquemas semânticos corporificados, ou seja, as construções são corporificadas segundo os modelos de uso da linguagem; ao invés de serem apenas objetos descritivos, como o são na maioria das abordagens. Esses pares de forma-significado são componentes do conjunto hipotético de processos em que se engajam os usuários de uma língua: qualquer construção envolvida na produção ou na compreensão de um determinado enunciado deve, portanto, estar ancorada em algum evento de uso da língua; não devendo haver assim, construções comportamentais vazias.

Dessa forma, como afirmam Bergen e Chang (2013), a ECG considera que as pessoas usam a gramática de maneira significativa e funcional e o modelo caracteriza-se como uma teoria preditiva do uso da linguagem orientada empiricamente e implementada computacionalmente. Motivado pela questão de quais mecanismos cognitivos e neurais as pessoas mobilizam ao usar a linguagem, o formalismo da ECG desenvolve modelos computacionais nos quais esses mecanismos são descritos em termos de estruturas e processos: respectivamente, o que as pessoas parecem saber e como parecem aprender e usar esse conhecimento. Ao mesmo tempo, esses modelos são propostos com o objetivo de buscar conexões com seu substrato biológico para o desenvolvimento de uma teoria que ligue os níveis computacional e biológico.

Uma vez que a ECG visa dar conta dos mecanismos de uso da linguagem, as estruturas e processos propostas devem naturalmente ser restritos por evidências sobre como as pessoas realmente produzem e compreendem a linguagem. Semelhantemente a outras gramáticas de construções, a ECG também levanta a hipótese de que o

significado desempenha um papel importante na gramática, por isso, considera a pesquisa sobre o processamento humano do significado crucial. De acordo com Bergen e Chang (2013), a última década viu uma explosão de experimentos comportamentais e pesquisas com imagens cerebrais para a compreensão da linguagem, com foco no significado. Para os autores, uma descoberta fundamental é a de que as pessoas entendem os enunciados realizando simulações corporificadas por meio de percepções motoras e outros sistemas cerebrais que emulam experiências internas semelhantes às que teriam quando experimentam as cenas em si. Sendo assim, devido a essa descoberta, a ECG inclui simulação corporificada (BAILEY 1997; NARANAYAN 1997; FELDMAN, 2006) como um componente de uso da linguagem, onde palavras e outras construções são representadas como estruturas de conhecimento que conduzem experiências perceptivas e motoras que o usuário da linguagem simula.

Devido ao progresso substancial de estudos que tentam compreender como o significado funciona na compreensão da linguagem, a simulação corporificada recebe destaque por demonstrar que durante o processamento linguístico há a ativação de simulações corporificadas das cenas descritas. Como afirmam Bergen e Chang (2013), simulação corporificada é o envolvimento interno de sistemas cerebrais de modalidades específicas para criar ou recriar experiências não presentes fisicamente. A simulação corporificada pode ser visual quando partes do cérebro relacionadas ao sistema visual tornam-se ativas de maneira semelhante ao que ocorre quando estímulos externos visuais estivessem realmente presentes. A simulação corporificada também pode ser motora quando temos a ativação de partes do cérebro dedicadas ao controle motor simulando uma ação física sem realmente envolver movimentos musculares. As pessoas automaticamente (e principalmente inconscientemente) realizam a simulação motora durante uma variedade de comportamentos cognitivos, como imagens mentais intencionais, recordações e outras tarefas.

Essas simulações perceptivas codificam detalhes implícitos não mencionados explicitamente na linguagem. Por exemplo, as pessoas representam os objetos mencionados com as formas mais prováveis de suas orientações ou cores que teriam, dependendo do contexto (ZWAAN, *ET AL.* 2002; STANFIELD E ZWAAN 2001). Essas simulações podem adotar uma perspectiva particular, ou ponto de vista, a partir do qual os compreensores simulam a percepção do evento, e a pessoa gramatical (por exemplo, você vs. ele) afeta a perspectiva que é adotada (BRUNYÉ *ET AL.* 2009). Da mesma forma ocorre sobre coisas perceptíveis, quando as pessoas processam a linguagem em que há o

envolvimento de ações, elas se engajam em representações motoras dessas ações. Essas representações motoras são bastante específicas, já que até a parte do corpo usada para realizar a ação, por exemplo, a forma mais provável de como mão se comportaria ao realizar tal ação (BERGEN E WHEELER 2005; MASSON, BUB E WARREN, 2008).

Na ECG temos então, uma teoria em que as funções potenciais de simulação podem ser propostas e modeladas, de modo que previsões detalhadas sobre o processamento do significado podem ser produzidas. Mais especificamente, a ECG fornece os meios para explorar a hipótese de que as construções linguísticas servem como uma interface para a simulação corporificada.

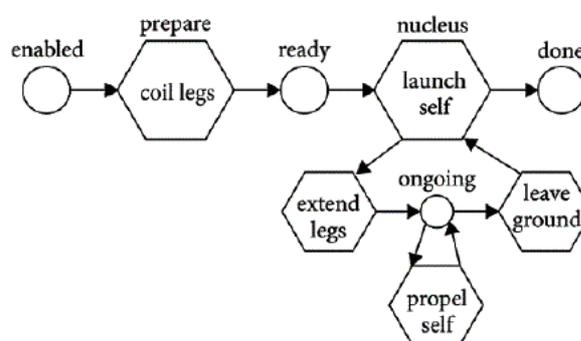
A ideia de que compreender a linguagem envolve a simulação corporificada exige um pensamento radical sobre o significado das palavras e construções linguísticas, bem como sobre o modo como esse significado deve ser representado. Tradicionalmente, a representação do significado foi compreendida conforme um conjunto de traços. Por exemplo, a palavra GATO apresenta um conjunto de traços que inclui [+felino, +doméstico]. Porém, na simulação corporificada, os significados são atribuídos pelos usuários através de ativações perceptivas ou motoras de representações correspondentes a entidades ou eventos presentes no processamento. As unidades linguísticas conduzem simulações corporificadas em vez de representações estáticas, formais e arbitrárias. Palavras e construções funcionam como interfaces que ativam e restringem simulações ricas em diferentes modalidades cognitivas.

Bergen e Chang (2013), seguindo a literatura linguística cognitiva, chamam as representações mentais categóricas que generalizam instâncias (como a forma ou generalizações de significado relevantes para a palavra gato) de *esquemas* (JOHNSON 1987; LAKOFF; JOHNSON 1980). E, seguindo abordagens gramaticais baseadas em construções, afirmam que toda a estrutura que liga tais esquemas é uma construção. As construções servem para conectar esquemas nos dois domínios, o da forma (fonológico, sintático) e do significado (conceitual, situacional).

A Figura 4, apresentada por Bergen e Chang (2013), fornece um exemplo de como as ações (neste caso, a ação de pular) podem ser representadas usando o “esquema de execução” (esquema-x). O esquema-x mostrado captura o fluxo de ativação por meio de um sistema motor responsável pela execução ou simulação de um salto. Primeiro, assumindo que todas as condições de habilitação são encontradas – por exemplo, o “pulador” está posicionado em uma superfície firme e tem energia suficiente para gastar –, ele se prepara movimentando suas pernas. Em seguida, vem a fase explosiva, em que

as pernas e o corpo são rapidamente estendidas e o “pulador” é impulsionado para cima a partir do solo, resultando em um período correspondente ao nó marcado por *Done*, em que o saltador no ar está sujeito à força da gravidade. Este esquema da ação de “pular” destina-se a abarcar uma variação considerável: o “pulador” pode ser um humano, um gato ou algum outro agente animado; a ação pode variar na força despendida, altura de lançamento atingida, direção do movimento ou orientação; e pode preceder uma série de outras ações como pousar no mesmo lugar, pousar em um local superior ou inferior à superfície, agarrando um objeto, voando, caindo etc.

Figura 4 – Esquema-x de pular



Fonte: Bergen e Chang (2013 p. 5)

Em termos de representação, a ECG representa o conhecimento linguístico também através de esquemas. Como exemplo temos um esquema conceitual simples que ilustra como a categoria de significado para a palavra *cat* é representada no formalismo da ECG. A representação (1) na Figura 5 inclui um nome do esquema (Cat), junto com uma notação indicando uma relação de herança, o *cat* é hipônimo de *animal*. O outro esquema, (2), representa a estrutura formal de *cat* e serve como indicador para representações fonológicas. Nesse esquema, temos a informação de que o esquema Cat-Form é um subcaso do esquema Word-Form mais geral, e traz ainda a característica ortográfica de *cat*. O esquema (3) de forma e significado está ligado por uma construção Cat-Cxn cuja forma é definida como uma instância do esquema Cat-Form, e cujo significado é definido como uma instância do esquema Cat. Cat-Cxn também é um subcaso de uma classe de construções semelhantes, denominada aqui de *Noun*. A construção *Noun*, não mostrada, emparelha de um lado uma forma não especificada e de outro um significado.

Figura 5 – Esquemas para a palavra *gato*

(1) schema Cat	(2) schema Cat-Form	(3) construction Cat-Cxn
subcase of Animal	subcase of Word-Form	subcase of Noun
	roles	form: Cat-Form
	orth ← “cat”	meaning: Cat

Fonte: Bergen e Chang (2013 p. 6)

Uma premissa comum às diferentes abordagens da Gramática de Construções é a continuidade entre léxico e sintaxe, por isso, a ECG propõe o mesmo tratamento para todas as unidades linguísticas, incluindo itens lexicais, bem como construções sintagmáticas maiores e noções gramaticais tradicionais, como mapeamentos entre forma e significado.

Ao contrário de palavras simples, no entanto, muitas construções exibem uma estrutura interna complexa com vários constituintes, cada um deles instancia um mapeamento entre forma-significado. Segundo Bergen e Chang (2013), em uma perspectiva baseada na simulação, isso significa que esses constituintes são particularmente bem adequados para parametrizar simulações corporificadas complexas envolvendo múltiplos esquemas conceituais variáveis.

Para ilustrar algumas características das construções gramaticais na ECG, Bergen e Chang (2013) recorrem à construção de estrutura argumental de Movimento Direcionado – *Directed Motion* – que licencia sentenças como *The horse raced past the barn* (O cavalo passou correndo pelo celeiro) e *The cat jumped down* (O gato pulou pra baixo.), em que um Movedor (designado por alguma Referring Expression) se move ao longo de um caminho (designado por uma Path expression) por meio de uma ação (designada pelo verbo). A construção *Directed Motion* especifica esses três constituintes e, na notação mostrada pelos autores, cada um recebe uma posição (r, v ou p) pela qual a construção pode se referir a eles. Cada um desses componentes também está restrito a um tipo de construção específico. Embora isso não seja mostrado aqui, as expressões de referência são aproximadamente análogas aos sintagmas nominais, enquanto as expressões de caminho expressam um caminho e incluem partículas e sintagmas preposicionais, como demonstra a Figura 6.

Figura 6 – Esquema Construção Movimento Direto

```

(6) construction DirectedMotion
subcase of ArgumentStructure
constituents
r: Referring-Expression
v: Verb
p: Path-Expression
form
ribefore vr
vibefore pr
meaning: Motion
mover ↔ rm
action ↔ vm
path ↔ pm

```

Fonte: Bergen e Chang (2013 p. 8)

A contraparte formal dessa construção, especificada na Figura 6, é mais complexa do que a de itens mais lexicais. Em seu significado, a construção denota um movimento, especificado pelo esquema de movimento. Na construção *Directed Motion*, os papéis do esquema de movimento são explicitamente ligados para os polos do significado dos constituintes de referência, verbo e caminho, respectivamente. Ou seja, o significado da expressão-referência, no caso do gato, será vinculado ao movedor do esquema de movimento como podemos observar na Figura 7

Figura 7 – Esquema Construção Movimento Direto

```

(7) schema Motion
roles
mover: Animate
action: Action
path: SPG
constraints
mover ↔ action.actor

```

Fonte: Bergen e Chang (2013 p. 8)

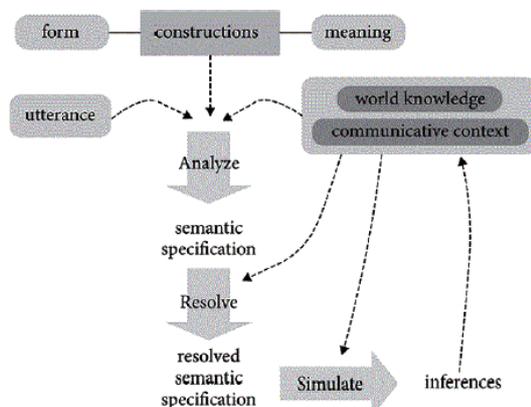
Em suma, as construções gramaticais na ECG são hipotetizadas para contribuir para a simulação corporificada através da ligação entre esquemas conceituais e seus constituintes, contribuindo diretamente para simular e impor restrições dentro dessa simulação. Como afirmam Bergen e Chang (2013), uma teoria da gramática que use o

formalismo da ECG deve incluir uma descrição de como a gramática restringe a simulação em termos mais amplos, contribuindo, assim, para que modelos implementados computacionalmente possam revelar empiricamente como construções específicas afetam a simulação corporificada.

Na Figura 8, encontra-se, de forma esquematizada, como funciona o processo de análise linguística conforme a ECG. Temos a análise estrutural, que ocorre pela identificação de quais construções são instanciadas por um determinado enunciado e como elas estão relacionadas a uma especificação semântica, *semspec*, bem como quais esquemas de significado são evocados e como eles estão relacionados. A Resolução contextual (*Resolve*), que se refere ao mapeamento de objetos e eventos na *semspec* para o contexto comunicativo (*communicative context*), fornece um *semspec* resolvido (*resolved semantic specification*). A simulação atua através da invocação de estruturas corporificadas para produzir inferências contextualmente apropriadas.

A Figura 8 também identifica várias informações trazidas para a compreensão de um enunciado: o conhecimento linguístico, o que os falantes sabem sobre a sua língua, ou seja, um conjunto de regras lexicais e gramaticais, associações estruturadas entre o conhecimento sobre a forma (como esquemas fonológicos) e conhecimento sobre o significado (como esquemas conceituais); o enunciado em si, o que os falantes percebem, ou seja, o sinal escrito, falado, assinado, gesticulado; o contexto comunicativo, como a situação física como o discurso precedente que produziram é apreendido; e o conhecimento do mundo, o que eles sabem sobre as entidades e eventos do mundo

Figura 8 – Visão geral das estruturas e processos envolvidos na compreensão de linguagem baseada em simulação



Fonte: Bergen e Chang (2013 p. 12)

Podemos observar que, quando consideramos aspectos semânticos, em implementações computacionais, através de cenas e esquemas, conseguimos tratar de forma mais satisfatória os fenômenos na língua, por isso, abordagens construcionais cognitivamente plausíveis fornecem subsídios para contribuir com os estudos de Compreensão de Língua Natural. Dessa forma, no capítulo 3 veremos como uma aplicação que utiliza os recursos da ECG pode ser interessante para o reconhecimento automático de construções. Na sequência deste capítulo conheceremos outra abordagem da Gramática de Construções, a Gramática de Construções Fluida.

2.2.2 Gramática de Construções Fluida

Apesar do grande interesse e crescente sucesso das abordagens construcionais em linguística empírica, linguística diacrônica, ensino de línguas, aquisição da linguagem e outros campos dos estudos da linguagem, ainda se observa uma certa ausência de formalismos para definir gramáticas de construções como também de uma implementação computacional acessível que operacionalize como construções podem ser usadas na análise ou produção de enunciados. Dessa forma, a Gramática de Construções Fluida tenta preencher essa lacuna.

O processamento computacional da linguagem exige um formalismo para representar léxicos e gramáticas, bem como algoritmos para análise semântica passando de um enunciado a uma representação do significado. Ainda não existe uma ampla aceitação de um formalismo para a GC e dessa implementação computacional robusta, mas alguns formalismos baseados na cognição, em oposição aos meramente computacionais, têm ganhado destaque, como é o caso da Gramática de Construções Fluida (FCG).

Segundo Steels (2011), a Gramática de Construções Fluida é um formalismo geral para definir e aplicar construções e pode ser usado para explorar diferentes abordagens teóricas, desde que a noção de construção seja reconhecida como uma organização fundamental. O autor argumenta que os léxicos e as gramática especificam relações bidirecionais entre forma e significado, passando pela semântica até as categorizações sintáticas, assim, diferentes construções expressam diferentes aspectos de tais mapeamentos bidirecionais.

Dentro da discussão sobre abordagens da GC e formalismos para se pensar um modelo de representação mais abrangente, Steels (2017) apresenta um conjunto de propriedades comuns a diferentes modelos de GC, importantes quando se pensa em uma

gramática baseada na cognição sendo utilizada em alguma implementação computacional.

Um desses princípios, segundo Steels (2017), diz respeito ao mapeamento bidirecional, em que diferentes estruturas – como a fonologia, a morfologia, a estrutura semântica, a estrutura sintagmática etc. – refletidas no enunciado são vistas como camadas horizontais e a cada camada corresponde uma representação de uma camada para a próxima. Assim, a gramática organiza os chamados esquemas de construções que podem abranger diferentes camadas.

Outro princípio apontado pelo autor é o da concordância. Diferentes teorias linguísticas que insistem em componentes estritamente modulares e em localidades hierárquicas rigorosas exigem que a informação relacionada à concordância seja movida para cima e para baixo de uma unidade para outra. Contudo, Steels (2017) afirma que não pode haver uma fronteira rigorosa entre as regras de cada camada. Quando se tem esquemas construcionais, eles devem ser capazes de contribuir diretamente para a construção de qualquer tipo de estrutura, e a representação dessas estruturas deve usar o mesmo formato para todas as perspectivas possíveis seja na fonética, na semântica, na sintaxe, ou na morfologia. Dessa forma, uma perspectiva não modular, além de lidar melhor com os aspectos não modulares da linguagem, pode conseguir um processamento mais eficiente não só porque todas as informações necessárias são acessíveis diretamente, mas também porque uma construção pode agrupar diferentes restrições necessárias e reverter ambiguidades.

Steels (2017) argumenta que a representação de construções através de esquemas apresenta-se mais eficiente pela possibilidade de serem aplicados de uma maneira flexível tendo algumas restrições relaxadas, havendo, assim, uma aplicação parcial das restrições permitindo a criação de uma interpretação parcial possível. Outro fator segundo Steels (2017) importante ao se tratar de construções pensando em um formalismo mais robusto é relativo à natureza da generalização das construções. Uma vez que construções capturam generalizações e essas estão presentes em diferentes esquemas de construções, o número de construções lexicais é minimizado já que algumas das informações funcionais ou de valência que estão, geralmente, contidas em entradas lexicais são movidas para construções gramaticais.

O grande desafio dos diferentes formalismos em GC, segundo Steels (2010), concentra-se em três principais pontos, que são: economia na representação, havendo uma única representação para todos os esquemas de construção; mesma representação para a

produção e compreensão, não sendo necessário criar mecanismos diferentes de uma para a outra; compreensão preditiva, poder de simular o que vai acontecer, sendo essa capacidade crucial na aprendizagem. Seguindo esse raciocínio, existe um outro aspecto merece ser destacado dentro dos princípios apontados por Steels (2017): o conceito de rede de construções. Construções não são entidades isoladas, são nós em várias redes que atendem a uma variedade de propósitos, como otimizar o processamento e simplificar a aprendizagem, além de representar diferentes particularidades das construções.

Conforme Steels (2017), as redes de construções possibilitam identificar uma generalização abstrata, assim como muitas variações com detalhes cada vez mais idiossincráticos que também devem ser explicitamente representadas, uma vez que as relações entre redes de construções são cruciais para a aprendizagem. Assim, uma nova construção não precisa ser aprendida, mas pode começar a existir a partir de seus semelhantes com variações apropriadas para a situação proposta.

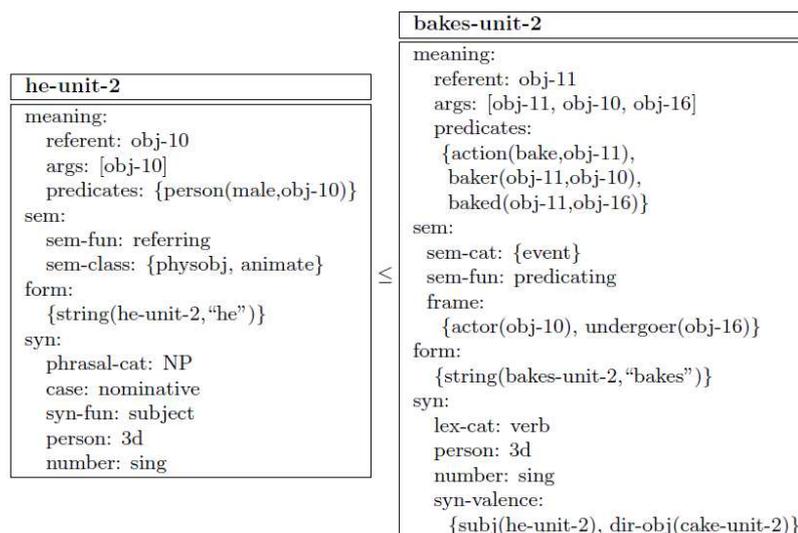
Diante de tantos desafios apontados para um formalismo mais adequado às questões postas acima, a FCG usa uma estrutura de traços para representar estruturas transitórias que representam informações de qualquer perspectiva de forma declarativa e explícita. Na FCG, as unidades nas estruturas transitórias correspondem a palavras ou sintagmas. Cada unidade possui um nome. O nome deve ser único dentro do contexto de uma estrutura transitória particular, mas o mesmo nome pode ser usado em diferentes estruturas transitórias. Cada unidade possui um conjunto de traços, que é um conjunto de atributos e valores. Os traços fornecem informações a partir de diferentes perspectivas linguísticas. Aspectos do significado ou da forma de um enunciado são representados como traços e espalhados por diferentes unidades e coletados no final do processamento para fornecer o significado total ou a forma total do enunciado.

A FCG se apresenta como um formalismo gramatical aberto em relação ao que se considera como unidades e ao que são atributos e valores possíveis dessas unidades. Segundo Steels (2010), depende da teoria linguística quais categorias e estruturas devem ser consideradas. Para o autor, as unidades em uma estrutura transitória formam um conjunto e cada unidade é acessível de forma independente dentro desse conjunto através do seu nome, o que proporciona grande eficiência e flexibilidade para lidar com línguas que não possuem uma estrutura sintagmática ou usam ordenação sequencial estrita.

Cada unidade em uma estrutura transitória é representada como uma caixa com o nome da unidade na parte superior e o conjunto de traços da unidade abaixo dela. Os traços semanticamente orientados, chamados de polo semântico, são geralmente escritos

acima dos traços orientados sintaticamente, chamados de polo sintático, mas na FCG não há ordenação fixa dos traços e nenhum conjunto definindo os traços que devem ocorrer obrigatoriamente. Na Figura 9 encontramos a representação de uma possível estrutura transitória do enunciado *he bakes*.

Figura 9 – Possível estrutura transitória para o enunciado *he bakes*



Fonte: Steels (2017, p.12)

Como podemos observar na Figura 9, existem duas unidades, uma para cada palavra. As unidades são chamadas *he-unit-2* e *bakes-unit-2*. Vemos aqui traços orientados para o polo do significado geral da construção, os quais dizem respeito ao atributo *meaning* e trazem informações como *referent*, especificando a correspondência dessa palavra dentro do formalismo utilizado. O atributo *args* diz respeito à correspondência dos argumentos exigidos pela palavra dentro da FCG, uma vez que em *predicates*, temos a valência semântica da palavra. Outros traços orientados para propriedades semânticas aparecem no atributo *sem* em que temos *sem-cat* – categorias semânticas –, *sem-fun* – função semântica – e *frame* – cena de valência semântica. Também vemos traços orientados para a forma e a sintaxe, como *form* que traz a informação do polo formal da construção que especifica propriedades sintáticas em *syn* como *syn-cat* – categorias sintáticas –, *syn-valence* – valência sintática – e *syn-fun* – função sintática. A unidade *he-unit-2* similarmente representa vários aspectos de uma ocorrência da palavra *he* incluindo informações sobre os traços semânticos, formais e sintáticos.

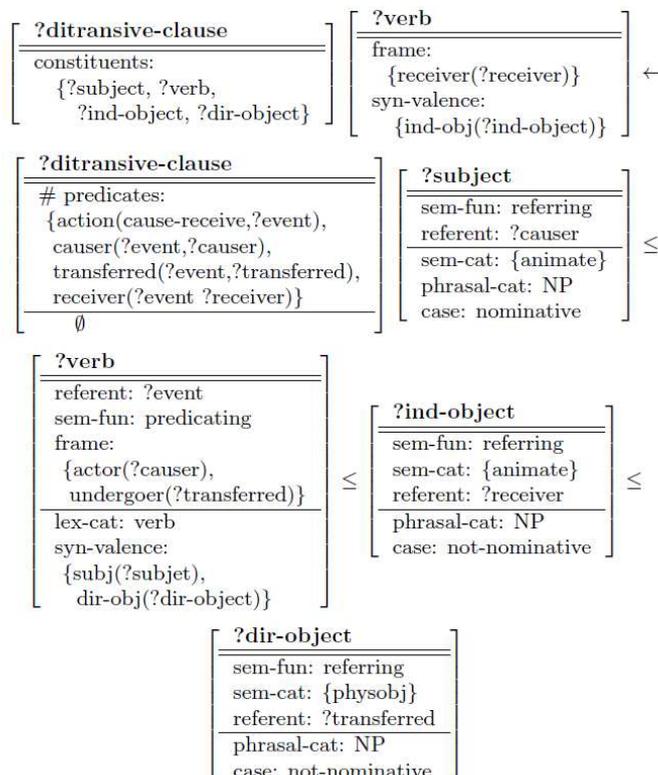
As unidades normalmente têm um traço de forma que coleta informações em termos de um conjunto de predicados. Um desses predicados é o *string* composto por dois argumentos, a unidade e a própria *string*, como em *form*: {string (he-unit-2, “he”)}, em que *he-unit-2* é o nome de uma unidade. No formalismo da FCG, as relações de ordenação não são representadas explicitamente, mas um sinal de <- é colocado entre as unidades relevantes para o termo precedente e um sinal \leq , para a relação de encontro entre as unidades. As unidades em uma estrutura transitória formam um conjunto não ordenado, por isso a gramática é fluida, já que várias relações podem ser impostas às unidades e elas são explicitamente representadas usando atributos e valores. Por exemplo, a estrutura sintagmática hierárquica pode ser representada como constituintes de um traço que contém os seus valores, ou pode-se representar unidades por relação de dependência, tratando os traços como dependentes, conforme o modelo de gramática que for escolhido. Algumas abordagens para a estrutura sintagmática apresentam uma representação das unidades em árvores, mas outras, como as baseadas em dependências, representam as unidades em um grafo. Dessa forma, como afirma Steels (2017), a FCG é altamente dinâmica, pois não restringe a representação a uma única perspectiva gramatical, já que é difícil prever em muitos casos como as unidades se relacionam entre si, portanto não seria possível restringir uma representação a uma árvore ou um grafo.

No formalismo da FCG, as construções são representadas através de esquemas de construções. Na Figura 10, temos um exemplo da representação do esquema correspondente a uma construção Bitransitiva que licencia construtos como “Pedro deu flores a Maria” e “Pedro assou um bolo para Maria”. Nessa representação, temos, na parte superior, duas caixas que indicam a estrutura básica dessa construção nos dando a informação de sua constituição bem como um possível significado conforme seu predicador. Abaixo, temos as estruturas transitórias que se encaixam nesse modelo, adicionando informações necessárias para o processamento. Esse encaixamento pode ser observado pelo uso do sinal \leq . Informações sobre as características do sujeito como estrutura sintagmática, papel semântico e caso são adicionadas. Da mesma forma o são informações sobre a valência verbal e conseqüentemente sobre o comportamento sintático-semântico dos objetos.

Podemos observar que, na estrutura básica a presença de um objeto indireto se encontra previsto, mas não no sentido de ser obrigatório, uma vez que esse formalismo, por se apresentar como genérico e econômico, tem que lidar com sentenças “*she gave him a book*” (ela deu um livro para ele), em que o receptor já está incluído na

valência sintática e semântica do verbo, mas também tem que lidar com sentenças como “*she baked him a cake*” (ela fez um bolo para ele), em que o receptor é adicionado pela construção.

Figura 10 – Construção Bitransitiva na FCG



Fonte: Steels (2017, p.28)

Apesar de a FCG representar um formalismo com grande potencial, que vem sendo aplicado em diferentes tarefas de processamento construcional (VAN EECKE E BEULS, 2018; VAN EECKE, 2017; BEULS E STEELS, 2017), o custo de implementação do Constructicon da FN-Br nesse modelo seria muito elevado visto que isso implicaria em reescrever todos os frames que já existem na FN-Br. Isso ocorre porque, conforme a classificação proposta por Boas, Lyngfelt e Torrent (2019), o Constructicon da FN-Br é profundamente integrado à rede de frames.

Nesse sentido, a proposta Gramática Construções de Berkeley se apresenta como mais adequada para nossa pesquisa, uma vez que temos um modelo de gramática baseado em unificação que possibilita em uma implementação computacional a integração de uma rede de frames a uma rede de construções.

Na próxima seção apresentaremos a Gramática Construções de Berkeley e posteriormente as implementações computacionais originárias dela que estão sendo desenvolvidas pela FrameNet Brasil.

2.2.3 Gramática de Construções de Berkeley

Seguindo a mesma linha dos modelos baseados em unificação, a Gramática de Construções de Berkeley – *Berkeley Construction Grammar* (BCG) – foi a que inaugurou o modelo construcionista, propondo a noção de construções como o pareamento entre forma e significado já a partir de trabalhos de Fillmore (1968; 1977) ainda identificados com a Gramática de Casos. Amadurece posteriormente com a Semântica de *Frames* (FILLMORE, 1982; 1985) e aparece, de maneira seminal, no trabalho de Fillmore, Kay & O'Connor (1988).

Com o intuito de tratar das relações de dependência sintático-semânticas entre os constituintes que estão intuitivamente relacionados ao modo como as pessoas conceptualizam as experiências e eventos em suas línguas por meio das sentenças, Fillmore (1968) postulou o *Case frame*, que buscava ligar essas descrições de situações a representações sintáticas subjacentes para encontrar um caminho para uma sintaxe baseada em signos e não só na forma, uma vez que esta manifesta linguisticamente uma cena cognitiva. Dessa forma, a Gramática de Casos forneceu as bases para o desenvolvimento da Semântica de *Frames* (FILLMORE, 1982), pois o princípio norteador de ambas as teorias é o de que “os significados são relativizados às cenas” (FILLMORE, 1977, p. 59).

Assim, a Semântica de *Frames* elege como tarefa tratar dessas relações combinatórias (sintático-semânticas) previstas pelo modelo gramatical de Fillmore (1968), colocando a semântica no centro do debate.

A noção de *frame*, que emerge na semântica linguística pela insuficiência da descrição do significado lexical em termos de uma lista de condições necessárias e suficientes, é entendida como uma estrutura conceptual complexa, organizada de tal modo que, para compreender qualquer de suas partes é imprescindível o conhecimento do todo, e, para a total compreensão dessa cena, é necessário saber os papéis semânticos que cada constituinte assume em relação à unidade lexical evocadora do *frame* (FILLMORE, 1982).

Temos como exemplo os itens lexicais comprar e vender, que ativam a mesma cena, ou seja, o mesmo *frame* de evento comercial, mas sob perspectivas diferentes. Enquanto a primeira perfila a perspectiva do comprador, a segunda perfila a do vendedor.

Assim, temos que o conhecimento linguístico do falante é estruturado sobre o *frame* que inclui o conhecimento sobre as diversas formas de realizar as diferentes partes de uma cena na forma de um enunciado. A ideia é que sempre que entendemos uma expressão linguística de qualquer tipo, temos, simultaneamente, uma cena (completa) de fundo e uma perspectiva sobre aquela cena que se dá através de seus participantes. Assim como os itens lexicais, as construções também podem evocar *frames*.

Diante desse histórico, a BCG é um modelo de gramática que integra os padrões sintáticos e semânticos das expressões linguísticas às noções da Semântica de *Frames* (FILLMORE, 1982; 1985). Ademais, a relação existente entre essa teoria e a Linguística Computacional fornece um formalismo baseado em unificação que é relevante para a proposta de desenvolvimento de recursos sintáticos, como Constructicons, e para tarefas de Compreensão de Língua Natural.

Fillmore, Kay & O'Connor (1988) apresentam construções como as regras que licenciam signos linguísticos baseados em outros signos linguísticos. As estruturas licenciadas por uma ou mais construções são chamadas de construtos. Qualquer construto pode ser representado pela junção das construções que o licenciam. A principal operação formal na BCG é a unificação entre os níveis de representação. A relação entre significado e significante é mapeada por um sistema baseado em traços em que as construções são representadas por Matrizes de Atributo Valor (AVM – Attribute Value Matrix), em um pareamento mediado por signos. As AVMs são combinadas de acordo com os valores atribuídos aos seus atributos e a coindexação mantém o controle da unificação de relações.

De acordo com Fillmore (2013), a BCG é uma gramática de estrutura de sintagmas que utiliza a representação de “caixas dentro de caixas” organizadas da esquerda para a direita em que cada informação sobre qualquer entidade linguística aparece como uma AVM associada às propriedades dos constituintes de tal entidade. Nas AVMs, os valores de cada atributo são compatíveis aos valores requeridos por cada posição sintagmática, já que, assim, é possível definir e especificar os tipos de entidades que podem ou devem estar em cada sintagma.

A Figura 11 exemplifica a representação da construção *DetHead*. Os colchetes são usados para identificar cada AVM. Dentro deles, estão contidos os atributos – *cat*, *max* etc. –, seguidos dos valores – *n*, *+* etc. – a eles atribuídos. Os atributos carregam as informações relacionadas aos signos linguísticos. Por exemplo, a maximalidade (*max*) possui valores polares *<+>* e *<->* e está relacionada com a possibilidade ou não de aquela

AVM constituir por si só um sintagma. Já categoria (cat) apresenta valores de nome, verbo, preposição, advérbio, adjetivo, determinante.

Figura 11 – Construção DetHead

$$\left[\begin{array}{l} \text{cat} \langle n \rangle \\ \text{max} \langle + \rangle \\ \left[\begin{array}{l} \text{cat} \langle d \rangle \\ \text{lxm} \langle \text{the} \rangle \end{array} \right] \left[\begin{array}{l} \text{cat} \langle n \rangle \\ \text{max} \langle - \rangle \end{array} \right] \end{array} \right]$$

Fonte: Fillmore (2013, p.3)

No exemplo da Figura 11, vemos que o colchete maior representa a construção *DetHead*, com a categoria de nome podendo constituir o núcleo de um sintagma próprio. As AVMs limitadas pelos conjuntos de colchetes paralelos representam, respectivamente, o determinante, com seus valores de lexema e categoria, e o nome, com os seus valores de categoria e maximalidade. A contraparte semântica da construção *DetHead* não aparece nessa representação, porém, de acordo com esse modelo, essa informação é explicitada pelo *frame* evocado por cada signo linguístico, sendo que os elementos que o constituem serão unificados aos signos que compõem a construção. Com a unificação entre a sintaxe e semântica, o objetivo desse modelo de gramática é mostrar as categorias gramaticais existentes no construto e que tipo de entidade linguística cada categoria representa em um determinado contexto, além de postular a formulação mais geral de regras que as colocam em uma construção.

O mapeamento semântico realizado na BCG para as construções ocorre via *frame*, assim, enquanto a contraparte sintática é representada via categorias e sintagmas, os frames são unificados a essas entidades conforme os Elementos de *Frame* relacionados a cada entidade, representando a contraparte semântica.

A noção de *frame*, que, conforme já se apontou acima, é outro dos desdobramentos da Gramática de Casos, surge na Linguística Cognitiva devido à incapacidade da descrição do significado lexical em termos de uma lista de condições necessárias e suficientes. É entendida como uma estrutura conceptual complexa, organizada de tal modo que, para compreender qualquer de suas partes, é imprescindível o conhecimento do todo, e, para a total compreensão dessa cena, é necessário saber os papéis semânticos que cada constituinte assume em relação à unidade lexical evocadora do frame

(FILLMORE, 1982). Nesse modelo, os participantes das cenas são denominados Elementos de Frame (EFs).

A noção de valência é outro aspecto relevante para a Gramática das Construções de Berkeley no tratamento construcional. Para Fillmore (2013), quando um lexema apresenta todos os valentes – semânticos e sintáticos –, especificados em termos de papel semântico e função gramatical, temos uma valência completa. A valência só requer um tratamento para além do lexical se as partes que a compõem não puderem ser definidas com base na valência de um item do léxico. Isso se deve ao fato de, se cada elemento que compõe a sentença puder ser explicado apenas com base no que se sabe sobre a valência do termo nuclear, logo, a valência da construção como um todo será derivada daquela de seu núcleo. Tome-se, por exemplo a sentença (8), fornecida por Fried e Östman (2004).

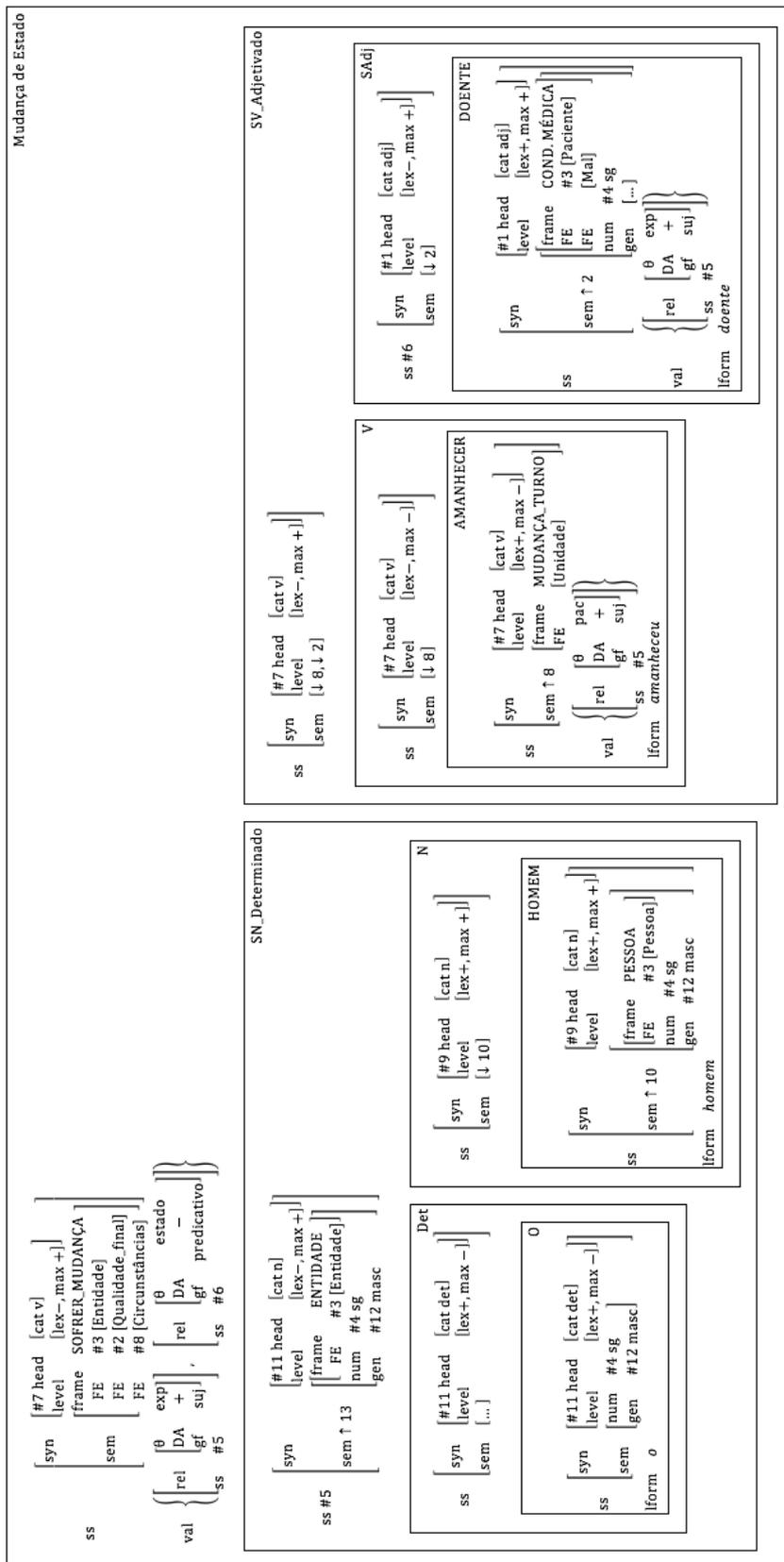
(8) Peter carried the bags.

Na sentença (8), observamos que o verbo *carry* sugere que dois participantes seriam necessários para que a sentença estivesse minimamente completa. Para Fried e Östman (2004), uma valência possui assim um conjunto de papéis semânticos, cada um dos quais está diretamente ligado a um participante de um *frame* específico, por um lado, e a sua expressão morfossintática correspondente, por outro. Dessa forma, no caso do verbo *carry*, Peter deve estar unificado ao EF Carregador do frame *Carry* e deve ter a forma e o comportamento gramatical de sujeito. Por outro lado, *bags* deve estar unificado ao EF Carga e apresentar a forma gramatical de objeto direto.

Um exemplo de formalização de construções seguindo os princípios da BCG pode ser observado na Figura 12. Nesse exemplo, temos a formalização do construto *O homem amanheceu doente*, licenciado pela Construção de Mudança de Estado. Cada caixa na Figura 12 indica um componente na construção. Os valores de cada atributo podem ser expressos como binários (+/-), itens em uma lista fechada (v, det, subj, Whole...), ou não ser especificados (...).

Outras informações presentes na AVM indicam se um constituinte é singular ou plural (num), se é masculino ou feminino gen e a Valência (val) representa a valência do constituinte e/ou da construção como um todo. Cada valente é definido quanto à sua estrutura e quanto à relação (rel) que estabelece com o constituinte, tanto em termos das funções teta (θ) e gramatical (gf) a ele atribuídas, quanto ao fato de ser ou não um argumento semanticamente proeminente (DA +/-).

Figura 12 – Construto “O homem amanheceu doente” licenciado pela
 Construção de Mudança de Estado



Fonte: Criado pela autora

Os índices numéricos presentes na matriz indicam a unificação dos valores de cada atributo. Os números de 1 a 13 são, portanto, índices de unificação. Quando precedidos por uma seta para baixo, esses números indicam que algumas características semânticas dos constituintes remetem às de um signo filho. Quando precedidos por uma seta para cima, indicam que tais propriedades são projetadas para o nível hierárquico superior. Essas relações semânticas são indicadas pelo frame evocado pelo constituinte, o qual é definido por seu nome e pelos elementos de frame (EFs) que o compõem. O processo de unificação pode indicar dois aspectos importantes da construção. Por um lado, checa se os atributos sintático-semânticos dos constituintes estão em acordo. Assim, os índices #4 e #12, por exemplo, dão conta da necessária concordância nominal entre o sujeito e seu predicativo, bem como daquela entre o determinante e o nome. Por outro, pareia forma e sentido, ao mapear a semântica de cada elemento presente na construção. Assim, o índice #3 indica que a Entidade que sofre a mudança de estado é uma Pessoa, que pode ser enquadrada como o Paciente do frame de Condições_médicas. Já o índice #2 indica que o estado final de #3 é estar doente, enquanto #8 mostra que a Circunstância em que tal mudança se deu corresponde à passagem da noite para a manhã. Por fim, o atributo da forma lexical (lform) indica quais elementos da sentença que contêm o construto licenciado estão mapeados a quais constituintes.

Notamos que a apresentação desse construto nos informa que a construção tem um significado próprio e um *frame* diferente do evocado pelo verbo. A semântica do verbo evoca o frame Mudança_turno definido por uma mudança que ocorre nas partes dos dias, em que seu EF, Unidade, não é mapeado dentro de nenhum argumento, pois a valência mínima do verbo não exige nenhum argumento. Já a valência construcional nos permite identificar uma mudança de uma Entidade de uma situação inicial para uma situação final. Essa informação é encontrada apenas na descrição construcional, uma vez que a valência verbal não nos mostra a mudança no sujeito, nem se considerarmos o adjetivo *doente* como predicador. A construção evoca um frame próprio que traz a informação de que uma entidade, sujeito, que sofreu uma modificação após uma passagem de turno no decorrer do dia.

Apesar de pequenas divergências, os modelos de Gramática de Construções apresentados aqui estão todos relacionados com o desenvolvimento de uma gramática explícita, não derivacional, baseada na noção de construção gramatical como uma associação convencional de forma linguística e conteúdo, por isso, tanto o modelo de

Goldberg como os modelos baseados em unificação fornecem contribuições para a descrição e formalização de diferentes estruturas linguísticas.

Outro ponto convergente dentro dos modelos de Gramática de Construções diz respeito à organização das construções. As construções relacionam-se umas com as outras através de redes construcionais, e as relações que contribuem para integração dessas redes serão explicitadas na próxima seção.

2.3 REDES CONSTRUCIONAIS

A gramática de uma língua se constitui do conhecimento do falante sobre a língua, e as construções são unidades básicas do conhecimento linguístico, então, a gramática nada mais é do que um conjunto de todas as construções de uma língua, desde as mais específicas até as mais gerais.

A respeito dessa questão, Salomão (2009) afirma que essas unidades linguísticas não se apresentam de forma isolada ou em lista, mas sim relacionam-se entre si através de redes, conectadas por relações de herança. Assim, o conhecimento gramatical de uma língua corresponde à rede de construções dessa língua, que representa, de forma contínua, tanto suas generalizações como suas peculiaridades.

Assume-se, portanto, que este repertório de construções que constitui o conhecimento linguístico não seja desestruturado, mas organizado radialmente e por relações de herança (LAKOFF, 1987; GOLDBERG, 1995, CROFT, 2001; KAY & FILLMORE, 1999). As construções estão normalmente relacionadas entre si, através de redes construcionais que apresentam diferentes tipos de relações. Goldberg (1995) considera que cada construção tem um significado próprio e uma forma esquemática e convencional que é capaz de fornecer as mais abstratas generalizações para o licenciamento de outras construções.

No modelo de Goldberg são encontrados diferentes princípios psicológicos responsáveis pela organização linguística das construções em rede. Um desses princípios é o da Motivação Maximizada, que afirma que “se duas construções são sintaticamente relacionadas, podem ser motivadas semântica ou pragmaticamente” (GOLDBERG, 1995, p. 67). Outro princípio é o Princípio da Não Sinonímia, segundo o qual “se duas construções são sintaticamente distintas, tais construções devem ser distintas semântica ou pragmaticamente” (GOLDBERG, 1995, p. 67).

Em função do Princípio da Motivação Maximizada, as redes construcionais podem apresentar diferentes tipos de *links* de herança. Dentre as relações de herança detalhadamente descritas por Goldberg (1995) destacam-se os *links* polissêmicos, em que a partir de um sentido central outros sentidos podem ser estendidos a ele por herança. As especificações sintáticas do sentido central são herdadas por essas extensões; portanto, não precisamos assumir uma realização sintática para cada extensão.

Goldberg (1995) afirma que o padrão sintático bitransitivo está associado a uma família de sentidos relacionados, em vez de um único sentido abstrato. O seguinte padrão de polissemia foi observado:

- (9) Pedro deu flores a Maria. (X causar Y receber Z)
- (10) Pedro prometeu um carro a Maria. (X pretende causar Y receber Z)
- (11) Pedro assou um bolo para Maria

Os sentidos em (10) e (11) estabelecem *links* de herança polissêmica, a partir do sentido central em (9), que indica prototipicamente a transferência física. Nos casos de herança, os exemplos apresentam diferenças relacionadas ao fato de que a transferência é pretendida.

Ao se falar sobre redes construcionais, a relação de herança é imprescindível. Porém a relação de herança adotada por Goldberg difere da relação de Herança Completa proposta por Kay e Fillmore (1999). Para os autores, as construções herdeiras herdam em todos os níveis as características do item hierarquicamente superior. Essa relação ocorre, portanto, quando a Construção B contém a Construção A – hierarquicamente superior – acompanhada de outros elementos.

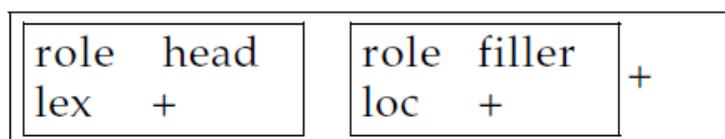
A respeito da relação de herança completa, Kay e Fillmore (1999) sugerem a existência de três construções nucleadas, conforme a sua relação com outras construções, a saber: (i) Núcleo_Complemento (HC), em que suas construções herdeiras consistem de um núcleo lexical seguido por um ou mais complementos; (ii) X_Núcleo (XH): compostas de um núcleo não necessariamente lexical precedido por um sintagma máximo, um especificador; e (iii) construções de modificação (Mod_X), em que uma palavra ou sintagma nuclear é precedido ou antecedido por um modificador. De acordo com Croft (2001), construções altamente esquemáticas, como é o caso das construções HC_Cxn, Mod_Cxn e XH_Cxn, são baseadas em padrões morfossintáticos que são postulados para justificar um tipo elevado de semelhança morfossintática e semântica.

Dessa forma, essas construções são usadas para definir outras construções através das relações de herança e fornecer padrões para a descrição linguística.

Para exemplificar a relação de herança completa, Kay e Fillmore (1999) apresentam a construção de Sintagma Verbal – VP (*verbal phrase*). A construção VP pertence a uma família de construções de sintagmas lexicalmente nucleados. De acordo com os autores, essa construção, herdeira da construção HC, especifica um núcleo lexical, o verbo, seguido ou não por um ou mais filhos preenchidos.

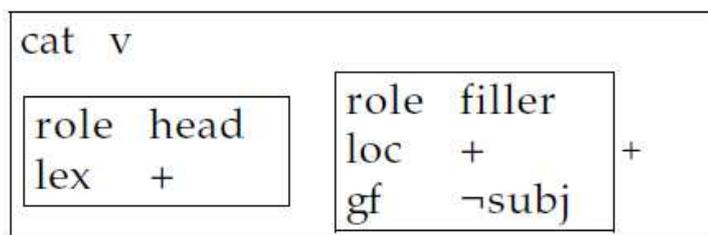
Como mostram a Figura 13 e a Figura 14, a construção HC fornece todas as características para a construção VP. A construção VP adiciona à informação que ela herda de HC apenas a categoria sintática de um núcleo verbal e o fato de que nenhum dos filhos preenchidos tem a função de sujeito.

Figura 13 – Construção HC



Fonte: Kay e Fillmore (1999, p.47)

Figura 14 – Construção VP

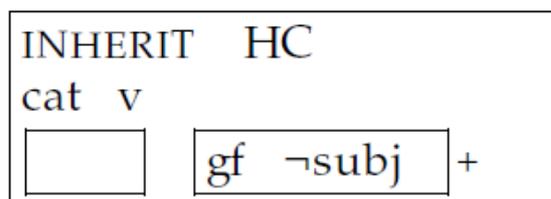


Fonte: Kay e Fillmore (1999, p.47)

Quando uma construção herda de outra construção, a primeira contém todas as informações da segunda. Como a construção VP herda da construção HC, todas as informações que aparecem na Figura 13 também aparecem na Figura 14.

Então podemos representar essa relação de herança na formulação da Figura 15. A notação “INHERIT HC” com o retângulo vazio indica que todas as propriedades da construção HC são compartilhadas pela construção VP, enquanto o retângulo preenchido apenas adiciona as informações novas.

Figura 15 – Construção VP representação alternativa



||Fonte: Kay e Fillmore (1999, p.47)

Nesse contexto, o modo completo de herança (Kay & Fillmore, 1999), o qual presume que a construção herdeira é um tipo mais específico da construção mãe, se mostra mais adequada para o nosso estudo, uma vez que um dos propósitos do *Constructicon* é apresentar o status de rede construcional. Essa relação fornece a possibilidade de armazenar informações de modo econômico e não redundante, tratando a herança pelo viés da cópia virtual de informação, isto é, a informação herdada é registrada somente na construção dominante, o que evita a redundância no armazenamento de dados.

Na próxima seção, o *Constructicon* e o *Lexicon* da FrameNet Brasil são apresentados como dois recursos computacionais que se utilizam do aporte teórico da Gramática de Construções e da Semântica de *Frames*.

2.4 IMPLEMENTAÇÕES COMPUTACIONAIS

Os pressupostos teóricos aqui apresentados da Gramática de Construções de *Berkeley* e da Semântica de *Frames* ultrapassaram o campo da linguística teórica para servirem de base em implementações computacionais, com o intuito de auxiliar em tarefas de Compreensão de Língua Natural. A FrameNet está envolvida em vários campos de pesquisa, que incluem da construção de recursos lexicais à identificação e catalogação de construções

A FrameNet começou em 1997 como um projeto de lexicografia computacional para o inglês, liderado por Charles J. Fillmore no *International Computer Science Institute*, em Berkeley, Califórnia (FILLMORE ET AL., 2003). Desde então, o recurso original vem expandindo sua cobertura de unidades lexicais para outros idiomas, como espanhol, alemão, sueco, japonês e português do Brasil. Nos projetos afiliados a

FrameNet de Berkeley, os *frames* e Elementos de *Frames* são expandidos para os outros idiomas, ou seja, é utilizada a estrutura definida para o inglês como ponto de partida para a descrição das unidades lexicais na língua de destino com adaptações e mudanças na estrutura original.

A FrameNet Brasil (FN-Br) é uma das iniciativas que adotaram a abordagem de expansão, ocupando-se no desenvolvimento de dois grandes recursos computacionais: um *lexicon* e um *constructicon*. Então, além de expandir a estrutura de frames e itens lexicais da Berkeley FrameNet para o português brasileiro, a equipe da FrameNet Brasil também foca no desenvolvimento de um *Constructicon*, ou seja, um repositório de construções do Português Brasileiro.

2.4.1 O Lexicon

O Lexicon da FN-Br é expandido da Berkeley FrameNet e possui uma rede de *frames* em relação à qual as unidades lexicais (ULs) são definidas. Os *frames* também fornecem as categorias semânticas usadas na anotação de sentenças que exemplificam as ULs. Como *frames* são representações de estruturas cognitivas, definidas, de acordo com Fillmore (1982) como um sistema de conceitos relacionados de tal forma que, para entender qualquer um deles é necessário tem que entender toda a estrutura na qual ele se encaixa, as análises realizadas para as ULs, portanto, nos fornecem a descrição de suas propriedades de valência sintática e semântica.

Note-se, na Figura 16, como a FN-Br formaliza um *frame* e suas relações. O *frame* traz uma definição, a qual vem seguida dos EFs que o compõem, com suas respectivas definições. Os EFs são classificados em nucleares (*core*) e não-nucleares (*non-core*). Enquanto aqueles são necessários para que o *frame* se instancie, estes são opcionais.

que estes assumem em relação à UL, assinalando os nomes dos EFs nas cores correspondentes às que foram marcadas quando da criação dos *frames*. A camada de FG é usada para indicar as funções gramaticais de cada constituinte da sentença alvo. As funções gramaticais codificadas pela FN-Br são de externo (Ext), que atua como sujeito da sentença, objeto direto (DObj), objeto indireto (IndObj) e dependente (Dep), função que abarca casos de adjunção, além de aposto (Aposto) e núcleo (Núcleo), funções utilizadas quando as palavras alvo são nomes ou adjetivos (TORRENT & ELLSWORTH, 2013). Por fim, a camada de PT3 contempla a classificação do sintagma em que um dado constituinte se encontra, por exemplo, se verbal é marcado como SV, se nominal é marcado como SN.

Além do Lexicon, a FrameNet Brasil também apresenta um *Constructicon*: um repositório de construções gramaticais do Português Brasileiro. O *Constructicon* foi destinado principalmente a demonstrar a viabilidade prática e teórica de se construir um repositório de construções ao lado do *Lexicon*, com ferramentas e métodos compartilhados. O *Constructicon* foi inicialmente implementado seguindo a mesma metodologia do *Constructicon* da Berkeley FrameNet (FILLMORE ET AL., 2012), porém foram adicionados recursos diferentes com o objetivo de proporcionar um melhor tratamento para as entidades linguísticas. O *Constructicon* do PB será apresentado na próxima seção.

2.4.2 O Constructicon

Devido à constatação dos limites impostos pela anotação lexicográfica, a Berkeley FrameNet passa a desenvolver, em 2008, em paralelo ao já estruturado *Lexicon*, um *Constructicon*, com o intuito de representar computacionalmente determinadas estruturas linguísticas não processáveis lexicograficamente (FILLMORE, 2008; FILLMORE ET AL., 2012).

Seguindo o proposto na BCG (FILLMORE, 2013), o *Constructicon* abarca o conhecimento linguístico que excede a valência simples de palavras simples (FILLMORE, 2008). De modo mais específico, descreve construções em termos de suas propriedades gramaticais e seu potencial semântico. A iniciativa de criação de um modelo computacional de construções foi seguida também pela FrameNet Brasil que, desde 2010, desenvolve um *Constructicon* para o PB (TORRENT ET AL. 2014; SILVA ET AL., 2017).

Da mesma forma que acontece para unidades lexicais, o *Constructicon* também realiza anotações, porém enquanto o método lexicográfico representa apenas a UL e os seus elementos de acordo com seu padrão de valência, o método construcional mapeia uma construção formalmente, unificando-a, quando relevante, a um *frame* específico.

Como pode ser visto na Figura 18, as construções são definidas de modo semelhante aos *frames*. Elas também apresentam uma definição e um conjunto dos Elementos de Construção (ECs) que representam os seus constituintes. Na construção Ergativa, os ECs são um Sujeito e um Predicado, que, por sua vez, compreende apenas um núcleo verbal. No campo Relações, registram-se as informações relacionadas ao *frame* evocado, *Ser_Afetado*, no qual um paciente sofre a ação de um evento, como também a sua relação de herança com a construção *Sujeito_Predicado*.

Figura 18 – Construção Ergativa

Ergativa [cxn_ergative]	
Definição	Tipo de construção SN_V. Essa construção exibe um argumento interno Sujeito e um Predicado .
Exemplo(s)	
Elementos da Construção	<p>Predicado [ce_predicate] O Predicado é composto por um SV.</p> <p>Sujeito [ce_subject] O Sujeito é um SN.</p>
Relações	<p>Evoca Ser_afetado</p> <p>Herda de Sujeito_predicado</p>

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Portanto, para desenvolver o *Constructicon* de acordo com os princípios da Semântica de *Frames* e da Gramática de Construções, foi necessário: (i) fornecer um meio de ligar construções a *frames* da mesma maneira que ULs estão conectadas a *frames*, (ii) modelar relações entre construções, de modo que o recurso possa fornecer como resultado uma rede de construções, e não uma lista delas.

Na Berkeley FN, apenas as relações entre *frames* e ULs e entre *frames* e outros *frames* são modeladas. Já no *Constructicon* do Inglês (FILLMORE ET AL. 2012), a conexão entre *frames* e construções, quando uma construção evoca um *frame*, é realizada na descrição textual da construção, mas não modelada no banco de dados. O mesmo acontece para os casos das relações de herança entre construções, quando uma construção é definida como um subtipo mais específico de outra construção, esta informação é fornecida apenas na descrição, e não no modelo.

Para tanto, no *Constructicon* do PB, fez-se necessário representar a Evocação de *frames* e a Herança de construcional adequadamente (LAGE, 2018). Para isso, consideraram-se as construções conforme sua semelhança com as ULs, na medida em que elas são um pareamento de uma forma e um significado, como também sua semelhança com os *frames*, uma vez que também apresentam estrutura interna.

A primeira relação, Herança, conecta uma construção à sua construção mãe. Conforme explicitado anteriormente, ela é modelada seguindo Kay e Fillmore (1999). Assim, a construção filha contém todas as informações da construção mãe e mais outras informações adicionais. Como um exemplo, considere as construções Sujeito_Predicado e Transitiva_Direta_Ativa, correspondentes à Figura 19 e à Figura 20

Figura 19 – Construção Sujeito_Predicado

Sujeito_predicado []	
Definição	A construção de Sujeito_predicado fornece um argumento externo (o Sujeito) a uma sentença que não o tem (o Predicado).
Exemplo(s)	
Elementos da Construção	<p>Predicado [pre] O Predicado é o predicado do Sujeito.</p> <p>Sujeito [sub] O Sujeito é identificado como o argumento externo do Predicado.</p>
Relações	<p>Herdado por Argumento cindido, Ergativa, Intransitiva, Mudança_de_estado, Predicativa_locativa, Predicativa_nominal, Transitiva Direta Ativa, Transitiva_indireta</p> <p>Herda de X_núcleo</p>

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 20 – Construção Transitiva Direta Ativa

Transitiva Direta Ativa [cxn_act_trans_dir]	
Definição	Tipo de construção SN_V_SN. Essa construção exibe um argumento externo Sujeito e um Predicado . O argumento interno é um objeto direto.
Exemplo(s)	
Elementos da Construção	<p>Predicado [ce_predicate_act_trans_dir] O Predicado é composto de um SV e um SN objeto direto.</p> <p>Sujeito [ce_subject_act_trans_dir] O sujeito é um SN.</p>
Relações	<p>Evoca Ação_transitiva</p> <p>Herdado por Objeto_interdito</p> <p>Herda de Sujeito_predicado</p>

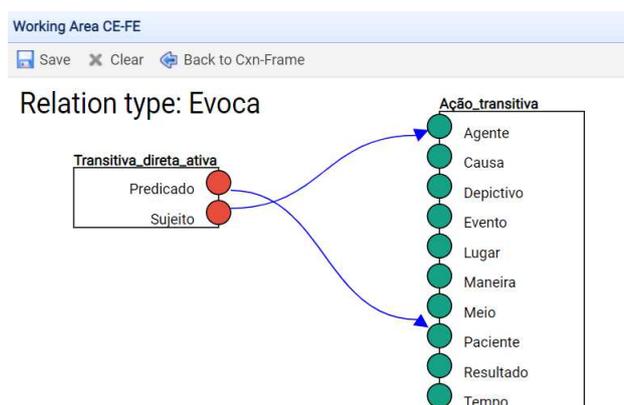
Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

A construção *Sujeito_Predicado* não apresenta uma instanciação própria. É uma construção abstrata responsável por um padrão sintático básico. Nessa construção, não há especificação sobre o Predicado, isto é, ele pode ser composto de um verbo, um verbo e um sintagma nominal, ou um verbo, um sintagma nominal e um sintagma preposicional, dependendo da transitividade do Predicado. Por outro lado, a construção *Transitiva Direta Ativa* nos informa que seu predicado deve compreender um Verbo e um Sintagma Nominal. Assim temos que toda a informação da construção de *Sujeito_Predicado* é compartilhada via herança para a Construção *Transitiva Direta Ativa*.

A segunda relação, *Evocação*, conecta uma construção a um ou mais *frames* que ela pode evocar. Diferentemente da *Herança*, a relação de *Evocação* não implica que o *frame* evocado seja “mãe” da construção que o evoca, isto é, essa não necessita conter toda a estrutura de seu esquema “mãe”. Assim, o *frame* evocado pela construção é evidenciado através do mapeamento entre os ECs da construção e os EFs do *frame* que ela evoca.

A conexão é proposta entre os elementos de composição em cada estrutura. Esse é o caso da construção *Transitiva Direta Ativa*, que evoca o *frame* *Ação Transitiva*. As relações CE para FE são colocadas através de uma interface gráfica, como mostrado na Figura 21.

Figura 21 – Relação Evocação frame *Ação_Transitiva*



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

A construção Transitiva Direta Ativa é uma construção que apresentam dois argumentos sintáticos e dois argumentos semânticos, um com função Agente e outro com função Paciente como em (13).

(13) O bandido queimou as provas.

O *frame* através dos EFs círculos verdes, estão relacionados à construção pela relação de Evocação, seus EFs estão unificados aos ECs – Sujeito e Predicado (círculos vermelhos) – da construção. Assim, os CEs correspondem ao Agente e ao Paciente no *frame* Ação_Transitiva.

A relação de Evocação contribui para integrar os bancos de dados do *Lexicon* e do *Constructicon*, uma vez que, sem tal relação, os recursos não interagiriam, e, juntamente com a relação de herança, tornam possível transformar a FrameNet Brasil em uma ampla rede de *frames* e construções interconectados

As construções no *Constructicon* da FN-Br também são definidas em termos de restrições. Na base de dados, uma restrição é uma relação entre um Elemento de *Frame* ou Elemento de Construção e qualquer Entidade, como um *Frame*, uma Construção ou um Elemento de Construção, que restringe a interpretação de um EF / EC.

Em um primeiro momento, foram criadas três restrições, a saber, a *CE>Constructicon*, a *CE_before* e a *CE_meets* (ALMEIDA, 2016). Posteriormente, outras três restrições foram elaboradas, quais sejam, *CE>Lexeme*, *CE>Frame* e *CE>Frame_family*, instanciando três níveis de restrições de preenchimento (LAGE, 2018). À medida que outras construções foram modeladas novas restrições se fizeram necessárias, como é o caso das restrições *CE > UDRelation*, *CE > UDPos* e *CE > UDFeature* (LAVIOLA, 2019) criadas de acordo com os princípios das *Universal Dependencies* (UD), assumindo que os ECs sejam definidos com base em características consideradas universais.

O Quadro 1 apresenta a definição de cada restrição, porém, no capítulo 5, serão demonstrados exemplos e aplicações das restrições.

Quadro 1 – Restrições no Constructicon da FrameNet Brasil

Tipo de restrição	Função
CE > Construction	Indica que o CE é licenciado por outra construção no Constructicon.
CE > Frame	Indica que todas as ULs que evocam um frame específico podem preencher um dado slot.
CE > Frame Family	Indica que todas as ULs que evocam uma família de frames possam preencher um dado slot.
CE > LU	Indica que apenas uma UL específica pode preencher um dado slot.
CE > Lexeme	Indica que apenas palavras específicas podem preencher um dado slot.
CE > Lemma	Indica que o CE é um lemma específico
CE > UDRelation	Indica que o CE apresenta uma relação sintática específica.
CE > UDPos	Indica que um CE pertence a uma classe de palavra específica.
CE > UDFeature	Indica traços como tempo, modo, aspecto.
CE > before	Indica que um CE deve preceder outro CE.
CE > after	Indica que um CE deve anteceder outro CE.
CE > meets	Indica que dois CEs devem ser adjacentes.
CE > Number	Indica singular e plural, mas foi incorporada em UDFeature.
CE > SemanticType of LU	Marca se um determinado CE precisa ou não ser definido como negativo ou positivo.
CE > Inheritance > CE	Indica que um CE é herdeiro de um outro CE em outra construção.
CE > Evokes > Concept	Utilizada para alinhamento multilíngue de construções.

Fonte: criado pela autora

O intuito da modelagem de restrições é proporcionar uma abordagem computacional a essas relações sintático-semânticas entre os constituintes que estão intuitivamente relacionadas ao modo como as pessoas conceptualizam as experiências e eventos em suas línguas através das sentenças.

Portanto, a FN-Br pode colaborar muito com estudos de Compreensão de Língua Natural, uma vez que *Lexicon* e *Constructicon* conseguem abarcar aspectos peculiares que vão além da estrutura sintática com um tratamento construcional baseado em unificação e em *frames*. Assim, forma e significado recebem o mesmo cuidado.

3 LINGUÍSTICA COMPUTACIONAL E COMPREENSÃO DE LÍNGUA NATURAL

A Linguística Computacional, enquanto parte da Linguística, tem alcançado cada vez mais importância nas atividades de descrição e análise das línguas humanas, justamente por possibilitar o uso de informações de línguas naturais contidas em diferentes bases de dados, ou seja, através da Linguística Computacional, as manifestações da língua em uso ganharam novas possibilidades de análise. Por outro lado, o tratamento computacional das línguas naturais e o próprio processo de comunicação humano tem instigado os centros de tecnologia da linguagem humana a investirem significativos recursos técnicos, humanos e materiais na modelagem computacional da linguagem humana, entendida, aqui, como a criação de um modelo computacionalmente tratável do uso do léxico e da gramática de uma língua natural nas diversas situações comunicativas (MCSHANE, 2017).

Em trabalho fundador da área no Brasil, Dias da Silva (1996) propõe que o estudo em Linguística Computacional visa à implementação de sistemas computacionais em que a comunicação entre o homem e a máquina possa se dar através de uma língua natural, e não pelo uso de comandos de uma linguagem de programação qualquer. Sendo assim, esse estudo deve fornecer a base para a implementação de programas computacionais específicos que, de alguma forma, sejam capazes de lidar com objetos linguísticos. Dentro desse domínio, um vasto conjunto de tarefas está envolvido, como a tradução automática, tratamentos superficiais do texto e até análises profundas do ponto de vista sintático e semântico.

Devido ao fato de a Linguística Computacional ser intrinsecamente multidisciplinar, uma vez que seu surgimento ocorre da intersecção entre os estudos linguísticos e a área de Inteligência Artificial, o que temos nessa área são novas possibilidades de análises de manifestações concretas da língua em uso, como também novos recursos metodológicos em que informações sobre as línguas naturais presentes em diferentes bases de dados possam ser úteis em várias tarefas presentes nessa sociedade dominada pela tecnologia.

3.1 COMPREENSÃO DE LÍNGUA NATURAL

A Linguística Computacional se insere como uma subárea da Inteligência Artificial, e em muitos casos é considerada como sinônimo de NLP (*Natural Language Processing*), que consiste no desenvolvimento de modelos computacionais para a realização de tarefas que dependem de informações expressas em alguma língua natural. Sendo assim, o estudo do NLP deve fornecer a base para a implementação de programas computacionais específicos que, de alguma forma, sejam capazes de lidar com objetos linguísticos.

Diversos programas usam métodos de NLP para que a comunicação mediada por computadores e baseada na compreensão e produção de línguas seja a mais eficiente possível. Através dos corretores gramaticais, tradutores automáticos e programas de reconhecimento de voz, é possível perceber o impacto social e tecnológico que as línguas têm na interação do homem com as máquinas, assim, produtos trazidos pela análise linguística, como dicionários e gramáticas, ganham outra roupagem na Linguística Computacional como ontologias, redes semânticas e analisadores sintáticos e semânticos.

Por seu caráter multidisciplinar, as pesquisas em Linguística Computacional também envolvem aspectos cognitivos que se estruturam através da relação entre mente e linguagem. Nesse sentido, a modelagem dos processos cognitivos incluídos na produção e compreensão da língua nos permite criar e testar modelos que se apresentem como cognitivamente plausíveis.

Por meio de implementações computacionais, um modelo de língua pode fornecer subsídios a uma hipótese acerca do funcionamento cognitivo humano, ao, por exemplo, analisar corretamente a estrutura de uma frase e propor que a mente humana utiliza dos mesmos mecanismos, uma vez que esse modelo foi baseado em características cognitivas.

O estudo do NLP trouxe algumas reivindicações no sentido de que novos modelos vão surgindo na intenção não só de processar uma língua natural, mas de *entender*, *compreender* e *aprender* uma língua natural, tarefa que passa a ser nomeada de Compreensão de Linguagem Natural (ou Natural Language Understanding – NLU)

A NLU é entendida, em alguns contextos, como um subconjunto do NLP. Entretanto, os programas em NLP e NLU, conforme explica McShane (2017), possuem objetivos distintos e atuam de formas diferentes, embora apresentem a mesma intenção de contribuir para o desenvolvimento de tecnologias que irão facilitar a experiência humana e sejam, portanto, ambos de grande relevância.

Como afirma McShane (2017), pesquisas em Linguística Computacional, durante os últimos vinte e cinco anos, deixaram os trabalhos de NLU em segundo plano, enquanto eram desenvolvidos sistemas baseados em estatísticas e na manipulação de *corpora*, que, muitas vezes, utilizavam anotação manual para alimentar o processo de aprendizagem da máquina.

Esses trabalhos vêm produzindo uma série de aplicações bem-sucedidas, no que diz respeito, por exemplo, à extração de informações, geração automática de sentenças, tradução por máquina, entre outros. Contudo, conforme argumenta a autora, os sistemas de NLP tendem a apresentar uma “inabilidade de computar e registrar o significado (a base da compreensão de língua natural ou NLU)” (MCSHANE, 2017, p. 44). McShane (2017) discute cinco crenças relacionadas a NLU e NLP, que as colocam em oposição e competição, e apresenta argumentos que refutam tais crenças.

O que é importante destacar é que, para a autora, NLU e NLP são programas complementares. Ambos têm ambições e métodos diferentes, mas podem trabalhar em complementação para criar sistemas mais complexos e que deem conta de um maior número de fenômenos. Outro ponto que merece destaque é que, para ela, não se pode considerar retirar a NLU da cognição e isolá-la e, ainda assim, esperar que o significado seja analisado perfeitamente.

McShane (2017) reforça a tese de que sistemas de NLP e NLU, embora apresentem diferentes objetivos e utilizem métodos distintos, intencionam contribuir para o desenvolvimento de tecnologias que visem a ampliar a experiência humana. Contudo, se a própria linguagem humana tem sido frequentemente caracterizada como um sistema que relaciona forma e significado, é de se esperar que sistemas baseados em perspectivas puramente estatísticas falhem ou sejam incompletos, já que, no processamento de língua natural, negligenciam o aspecto do significado ou tratam tal aspecto de forma superficial.

Bender *et al.* (2015) demonstram que projetos em NLP e NLU se beneficiariam se utilizassem métodos que envolvessem representação semântica. Os autores explicam que a máquina não é capaz de captar o significado pragmático de certas expressões, visto que este é obtido a partir da análise de outros fatores, por exemplo, a intenção comunicativa do interlocutor.

As raízes do problema demonstrado por Bender e Koller (2020) já haviam sido apontadas, em outros termos, em obra fundadora do campo da NLU. A partir das considerações de Allen (1985), identificamos que a principal diferença entre NLP e NLU se encontra no fato de que a primeira abordagem é de cunho estruturalista e seu foco se

concentra nas especificações formais das estruturas linguísticas, tanto das regras de construção quanto das restrições de constituição de estruturas possíveis. Já a segunda se preocupa com aspectos cognitivos e psicolinguísticos, levando em consideração não só as representações das estruturas linguísticas, mas também as dos processos linguísticos realizados na linguagem.

Dessa forma, a NLU visa a mapear enunciados em algum tipo de representação semântica que modela seu significado de maneira que um computador possa processá-los e entendê-los (ALLEN, 1985).

Existem duas motivações principais para o desenvolvimento da NLU, segundo Allen (1985): uma é a motivação tecnológica, uma vez que seu objetivo é construir sistemas computacionais inteligentes; a outra motivação é linguística (ou da ciência cognitiva), pois busca-se compreender melhor como os seres humanos processam e compreendem a língua, ou seja, como os humanos usam a língua.

Portanto, as pesquisas em NLU procuram trazer para o debate em Linguística Computacional a representação semântica de modo que um programa de Inteligência Artificial desenvolva a compreensão de uma língua natural tendo como objetivo real criar mecanismos para entender o significado do texto, sendo esse oral ou escrito, a fim de gerar uma resposta compreensível. Sendo assim, esse tipo de programa deve contemplar os aspectos da linguagem e da cognição humana, bem como conhecimento de mundo e habilidades de raciocínio que atuam na produção e na compreensão de estruturas linguísticas (ALLEN, 1985).

De modo a integrar esses aspectos no desenvolvimento de programas computacionais baseados em língua natural, Allen (1985) já propunha uma abordagem de NLU dividida em três subáreas: estrutura linguística, conhecimento de mundo e contexto situacional que remetem aos processos formais – fonológicos, morfológicos e sintáticos –, semânticos e pragmáticos, respectivamente. Assim, o processo sintático envolve análise de sentenças em língua natural e produz a representação dessas estruturas sintáticas. O sistema então usa essas estruturas como entrada na fase de interpretação semântica, a qual produz uma forma lógica que, por conseguinte, é utilizada como entrada para a fase de representação contextual, a qual desenvolve a representação final das sentenças.

O que merece ser pontuado ainda sobre a perspectiva da NLU é que, mesmo para tarefas cujos resultados sejam focados em aspectos formais, tais como *parsing*, por exemplo, podem-se obter resultados melhores na representação sintática quando

informação semântica é trazida para o sistema, confirmando a ideia de que, sendo a linguagem humana por si própria um sistema que relaciona forma e significado, uma representação computacional eficiente não pode evitar por completo essa associação.

Assim, os esforços da FN-Br têm sido canalizados para a constituição de um sistema de representação de construções que, naturalmente, une aspectos de forma e sentido dos fenômenos linguísticos. Assim, buscamos propor um modelo representacional capaz de apresentar diferentes construções de estrutura argumental no PB, considerando seus aspectos formais e funcionais, e implementá-lo através de um reconhecedor de construções.

As tarefas de reconhecimento automático de construções e *parsing* construcional já foram tentadas em trabalhos anteriores, os quais detalhamos a seguir.

3.2 TRABALHOS RELACIONADOS

Nesta seção apresentaremos trabalhos que se ocuparam em observar o processamento de construções. Dentro do domínio da NLU, um vasto conjunto de tarefas está envolvido nesse processo, que vão desde tratamentos de categorias específicas até análises profundas do ponto de vista sintático e semântico. Muitas dessas tarefas necessitam de tratamentos que incluam análises precisas que permitam fornecer, para uma dada sentença de entrada, as suas possibilidades de análise sintática e semântica, através das possibilidades combinatórias das classes de palavras contidas na sentença.

Atualmente, alguns trabalhos, amparados nos modelos de Gramática de Construções, têm se dedicado a propor análises automáticas de sentenças com o objetivo de apresentar como um modelo baseado em construções pode contribuir para uma análise mais abrangente e mais eficaz. Nas próximas seções veremos alguns trabalhos que utilizam essa estratégia.

3.2.1 ECG Analyzer

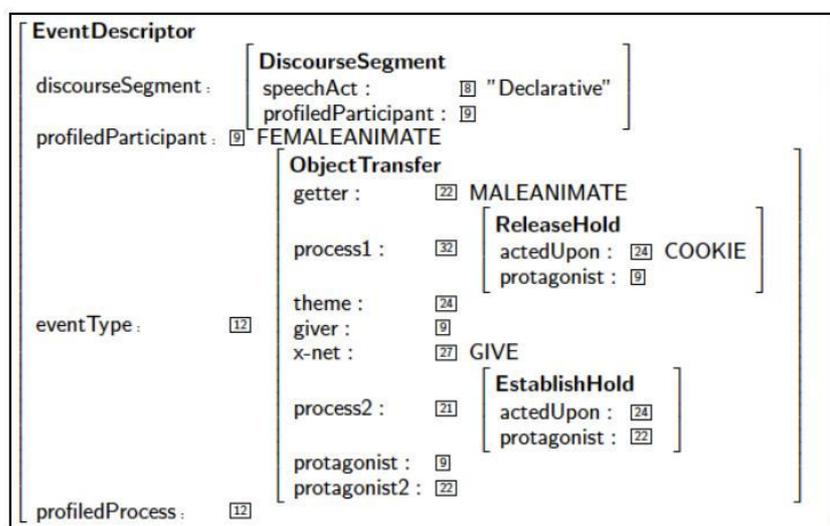
Dados os recursos oferecidos pela *Embodied Construction Grammar* (ECG), apresentada no capítulo 2, uma das implementações computacionais propostas para a identificação de construções é o *Constructional Analyzer* (BRYANT, 2008). Esse sistema é usado no contexto da ECG para a interpretação cognitivamente plausível de uma dada sentença. O *Analyzer* emprega a ECG como espinha dorsal: construções e estruturas

linguísticas e cognitivas, tais como esquemas imagéticos e *frames*, formam a base de cada interpretação desenvolvida pelo analisador. O *Analyzer* utiliza uma unidade computacional chamada de reconhecedor de construção, que é projetada para verificar as restrições de forma e o significado das construções.

Em vez de um parser monolítico que usa regras gramaticais como entrada, o próprio *Analyzer* é um repertório de reconhecedores de construções que trabalham em conjunto para gerar a análise construcional. Devido ao emparelhamento inicial de forma e significado proporcionado por construções, esse analisador com acesso às informações semânticas é capaz de usar como *input* soluções heurísticas para os problemas comuns associados à análise da linguagem real.

Bryant (2008) afirma que a chave dessa abordagem e o que torna o sistema mais robusto é a ideia de que cada enunciado está tentando comunicar algum tipo de cena e, nessa cena, os participantes são parametrizados em *frames* e esquemas imagéticos na ECG, assumindo que uma melhor análise é aquela que melhor descreve a cena. Como exemplo, considere-se a especificação semântica gerada pelo *Analyzer* para a sentença *She gave him a cookie*, na Figura 22.

Figura 22 – Especificação no *Analyzer* da sentença *She gave him a cookie*



Fonte: Bryant (2008, p.145)

A Figura 22 mostra a especificação semântica para a sentença *She gave him a cookie*. Nela, o FEMALEANIMATE é o perfil de identificação do participante e está substituindo o pronome *She*. O tipo de evento é perfilado como uma transferência de

objeto. Esse tem o receptor definido como MALEANIMATE (*him*), o doador está definido como FEMALEANIMATE, e o tema definido como o *cookie*. Notamos que os números representam a unificação dos participantes de cada cena.

Feldman (2020) aponta que os esquemas individuais, que são a base da ECG, são definidos dentro de uma rede maior de esquemas e esses esquemas são definidos como composições para conceitos mais gerais. Na ECG as construções são formalizadas de forma que cada uma especifica a relação entre algum significado, representado usando esquemas, e a forma pela qual esse significado é expresso. As construções fornecem um meio para especificar como a linguagem compõe conceitos mais elementares em outros mais complexos. O significado de uma construção individual pode especificar como os significados de suas partes, por exemplo, palavras individuais, são combinados um com um outro.

Bryant (2008) apresenta uma série de análises com diferentes construções, mostrando como o *Analyzer* realiza as especificações semânticas de construções como transitivas, passivas, bitransitivas etc. Essas especificações são semelhantes ao exemplo mostrado na Figura 22, que exemplifica a construção Bitransitiva. Contudo, para testar o analisador de construções como um modelo de interpretação semântica profunda, Bryant (2008) mostra como ele pode ser usado para modelar a ambiguidade em construções relativas reduzidas através de testes de tempo de leitura produzidos por McRae et al. (1998).

O objetivo foi usar um modelo de interpretação baseado em construções, uma vez que, na gramática de construções, o léxico, itens, expressões idiomáticas e regras gerais de vinculação de argumentos sintático-semânticos são representados uniformemente. Esta uniformidade de representação torna um modelo baseado em construções capaz de fazer previsões sobre uma ampla gama de fenômenos linguísticos. Desse modo, o fato de que o analisador usa construções é uma característica importante do sistema porque permite que ele seja compatível com a modelagem de diferentes tipos de fenômenos.

No experimento realizado, Bryant (2008) defende que o analisador construcional foi projetado para ser um modelo cognitivamente plausível de interpretação e não um programa para prever o tempo de leitura, uma vez que já existem diferentes modelos que fazem isso. Para Bryant (2008), prever dados de tempo de leitura é apenas importante na medida em que fornece outra maneira de sugerir que o analisador estrutural é, de fato, cognitivamente plausível.

O experimento considerou os exemplos em (14) e (15).

- (14) The cop arrested by the detective was guilty.
(15) The crook arrested by the detective was guilty.

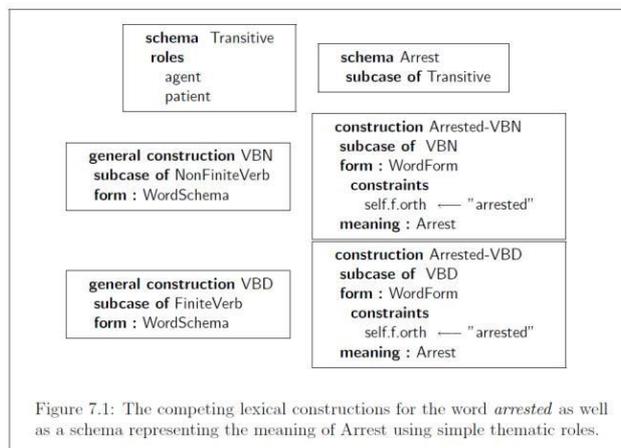
Ambas as sentenças usam a construção relativa reduzida (*arrested by the detective*), mas, o mais importante, os sujeitos diferem em termos de que é um agente prototípico para prender, *cop*, e o outro um paciente prototípico para ser preso, *crook*. Ademais, a forma *arrested* é ambígua, permitindo a leitura como participípio e tempo passado.

Bryant (2008) afirma que McRae, et al. (1998) testaram pares de sentenças como esse dentro de um paradigma de leitura individualizado. Eles descobriram que os tempos de leitura para as duas sentenças diferem dependendo se a sentença teve um sujeito agente prototípico para o verbo ou um sujeito paciente. A explicação básica para a diferença de tempo de leitura fornecida é que elas são uma consequência da violação da expectativa semântica: porque *arrested* é ambíguo entre o pretérito e o participípio e *cops* (policiais) são protótipos daquele que prende e não daquele que é preso, o tempo de leitura acaba sendo maior porque o participante do experimento precisa reorganizar a sintaxe da sentença à medida em que a lê. Na segunda frase, o cenário se inverte. A melhor análise sintática é, desde o princípio do processo de leitura, aquela em que o bandido preso pelo detetive era culpado.

Bryant (2008) ressalta que os resultados de McRae et al. (1998) mostram que as diferenças na semântica afetam a interpretação palavra por palavra em uma sentença. Além disso, coloca em questão qualquer hipótese que defenda uma sintaxe fortemente autônoma. E, finalmente, mostra que um modelo cognitivamente plausível de processamento de sentenças humanas deve incluir um modelo de semântica, e que o modelo semântico precisa condicionar as escolhas de processamento em um nível incremental.

Para que o *Analyzer* possa ser usado para modelar os dados de McRae et al. (1998), Bryant sugeriu que, primeiramente, deveria ser escrita uma gramática para as sentenças de teste e definida a sintaxe relevante e parâmetros semânticos para o modelo fatorado, passo semelhante ao adotado na modelagem das construções de estrutura argumental objeto desta tese.

Ao definir as construções lexicais Bryant fez um tagueamento em classes gramaticais e foram construídas pequenas hierarquias lexicais. A Figura 23 mostra o exemplo do verbo *arrested*.

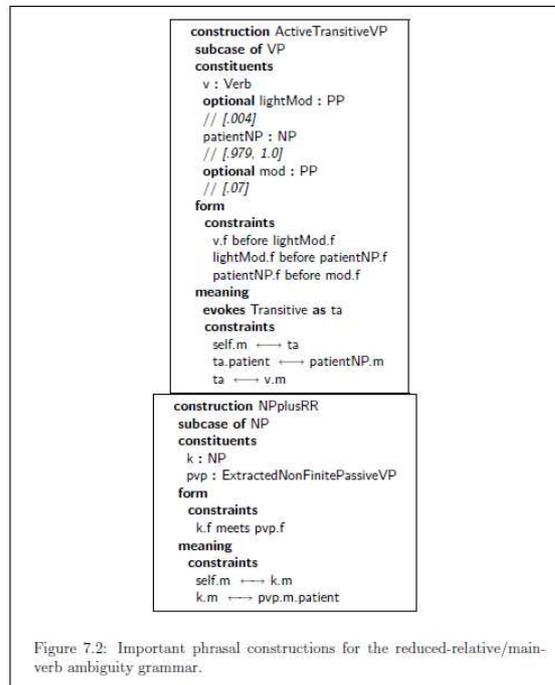
Figura 23 – Esquema de *ArrestedVBD*

Fonte: Bryant (2008, p.157)

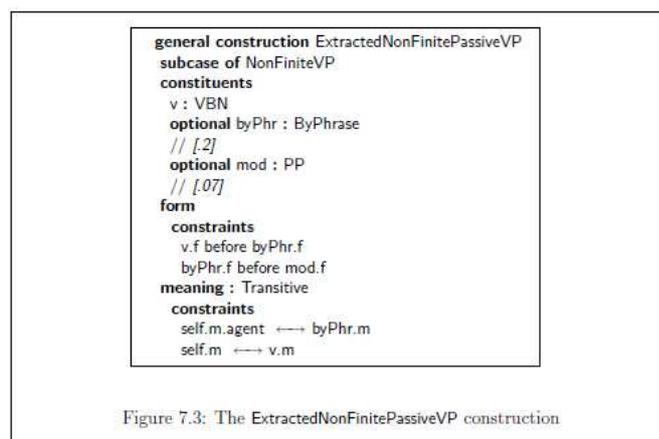
Na Figura 23 há a construção *Arrested-VBD*, que é correspondente a construção lexical *arrested* no pretérito, e a construção *Arrested-VBN* referente a classe gramatical para participípios passados, então temos a construção geral VBN como a raiz da hierarquia de participípio passado. Ambas as construções especificam que o esquema *Arrest* tem o significado transitivo do verbo *arrest*.

Três das construções importantes para a tarefa apresentada por Bryant (2008) são mostradas na Figura 24 e Figura 25. A Construção *ActiveTransitiveVP* é uma construção simples de sintagma verbal transitivo. Possui constituintes como um verbo *v*, um objeto paciente NP e dois modificadores PP opcionais *lightMod* e *mod*. O bloco que representa a contraparte formal restringe a ordem para ser *v*, *lightMod*, *patientNP*, *mod*. Já o bloco que representa a contraparte do significado evoca um esquema transitivo e o vincula a ambos os significados, o da construção e o do verbo. Isso limita os verbos que podem estar vinculados a essa construção, considerando apenas aqueles que são semanticamente transitivos. Adicionalmente, o papel do paciente no esquema transitivo é identificado com o significado de *patientNP*.

As probabilidades presentes no bloco de constituintes indicam a probabilidade de um constituinte ser expresso e, se expresso, o local em que ele se encontra. Para o constituinte *patientNP*, a probabilidade de ele ser expresso é estimada em 0,979 e, nesta gramática, é restrito ao local com peso 1,0. Os dois constituintes opcionais apenas têm sua probabilidade de serem expressos, uma vez que não podem ser expressos de forma não local.

Figura 24 – Construção *Active TransitiveVP*

Fonte: Bryant (2008, p.159)

Figura 25 – Construção *ExtractedNonFinitePassiveVP*

Fonte: Bryant (2008, p.160)

Também mostrada na Figura 24 é a construção relativa reduzida *NPplusRR* que une um NP constituinte *k* com um constituinte denominado *pvp*. O

ExtractedNonFinitePassiveVP é a cláusula relativa. O bloco que representa o significado de *NPplusRR* identifica o significado da construção com o sendo significado de *k* que, por sua vez, está relacionado com a função do paciente do constituinte *ExtractedNonFinitePassiveVP*. A construção mostrada na Figura 25 é a construção *ExtractedNonFinitePassiveVP*, que é um tipo de *NonFiniteVP* e tem um constituinte VBN, um constituinte opcional e um modificador PP opcional. O bloco que representa a forma requer que o verbo venha antes da *ByPhrase* e que a *ByPhrase* venha antes do modificador *mod*. O bloco do significado representa o esquema transitivo mostrado na Figura 23. O significado de *ByPhrase* está vinculado à função do agente, e o significado da própria construção está ligado ao significado do verbo que requer que apenas verbos com semântica transitiva sejam constituintes dessa construção.

Bryant (2008) afirma que o Analyzer usa três tipos de parâmetros: o parâmetro específico de localidade de cada constituinte, o parâmetro de preenchimento de constituintes e o parâmetro de função semântica.

O parâmetro de localidade é baseado na ideia de que uma gramática deve especificar a probabilidade com que um termo pode ser expresso ou omitido, e assumindo que o constituinte é expresso, a probabilidade de ele ser expresso localmente também deve ser especificada. Por exemplo, a Figura 24 mostra que a construção *ActiveTransitiveVP* tem um constituinte *patientNP* que pode ser omitido 2% do tempo.

Para os parâmetros de preenchimento de constituinte, Bryant (2008) faz uso de uma suposição de simplificação livre de contexto. Em vez de condicionar o parâmetro na construção e no constituinte, utilizou-se $Pr(\text{filler} = \lambda \mid \text{restrição de tipo constituinte})$ e, para as classes gramaticais, $Pr(\text{palavra} \mid \text{tag})$. Já para os sintagmas nominais, foi usada a estimativa simples, por exemplo, para a probabilidade de um dado pronome sobre o número de NPs que esse limita.

Em relação aos parâmetros semânticos, Bryant (2008) assume que as sentenças (14) e (15) de McRae et al. (1998) podem ser interpretados conforme os papéis semânticos de *cop* e *arrested* como em *Pr(probabilidade de o policial ser o agente de prisão \mid policial, prisão)*

Existem dois outros parâmetros semânticos nessas frases que afetam probabilidade de cada interpretação. Se o leitor tratar *arrested* como verbo principal, *by detective* deve ser interpretado como uma frase locativa. Se o leitor interpreta como o início de uma relativa reduzida, então *by detective* é interpretado como agente do esquema de prisão. Para modelar a diferença de tempo de leitura foram estimados parâmetros

semânticos do detetive enquanto um ponto de referência dentro de um esquema Trajetor_ponto-de-referência e um parâmetro semântico entre detetive e o papel do agente de prisão.

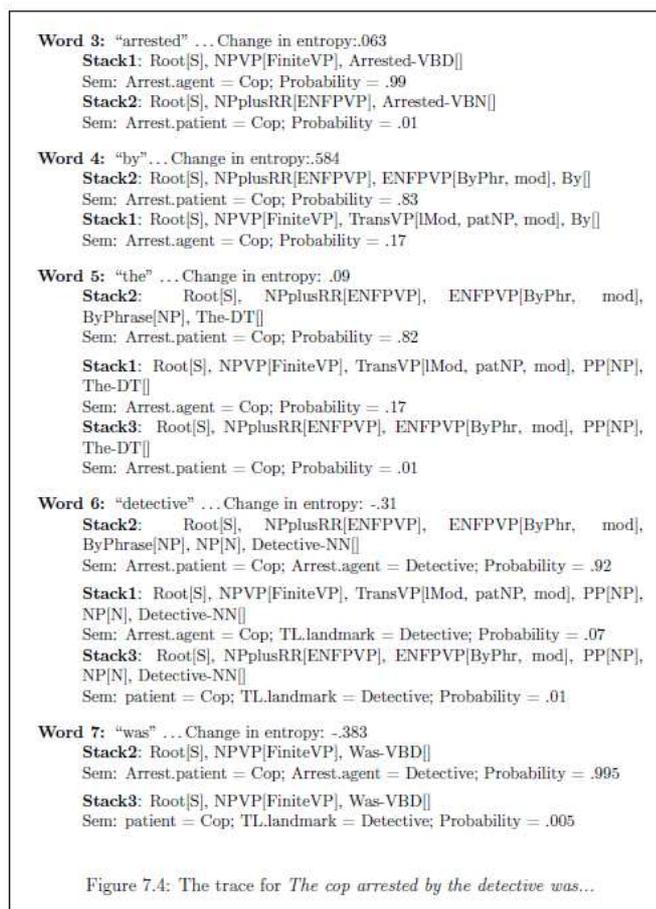
No experimento, para usar o *Analyzer* associado a atividades de tempo de leitura, Bryant (2008) utilizou a entropia, que consiste em medir a incerteza em uma probabilidade de distribuição dada uma variável aleatória X , a entropia é definida através da equação:

$$H(X) = \sum_{x \in X} \text{Pr}(x) \times \log \frac{1}{\text{Pr}(x)}$$

À medida que o analisador processa cada palavra do enunciado, ele mantém uma distribuição sobre interpretações. Esta distribuição estima a probabilidade condicional de cada interpretação. A equação soma cada resultado possível de x em X , escalando o log de $1 / \text{Pr}(x)$ pela probabilidade de x . Grandes valores de $H(X)$ significam alta incerteza e pequenos valores de $H(X)$ significa menos incerteza. Usando a equação, o sistema calcula a entropia condicional sobre as interpretações de cada palavra. A intuição é que um baixo valor de entropia significa que o analisador tem grandes chances de estar certo sobre qual interpretação é mais correta, enquanto uma alta entropia significa que muitos candidatos viáveis ainda permanecem sem ser interpretados.

As diferenças entre *cop* e *crook* mostram como o processamento difere entre os dois casos. Observemos a Figura 26, em que os traços usam os nomes de construção abreviados. “ENFPVP”, por exemplo, é a abreviação de *ExtractedNonFinitePassiveVP*. *lightMod* também é abreviado como “lMod”, a construção *ActiveTransitiveFiniteVP* é abreviada para “TransVP”, e o esquema TrajetorLandmark é abreviado para “TL”.

Considere a interpretação do *Analyzer* para *cop arrested by the detective was...* mostrado na Figura 26. Quando a palavra *arrested* é encontrada, duas listas são mantidas, uma para *arrested* como verbo principal e outra para *arrested* como relativa reduzida. A Stack1 vincula *Cop* ao papel do agente, Stack2 vincula *Cop* ao papel do paciente. É importante ressaltar que a principal interpretação do verbo é baseada tanto em bases sintáticas quanto semânticas, o que leva a poucas mudanças na entropia de *arrested*.

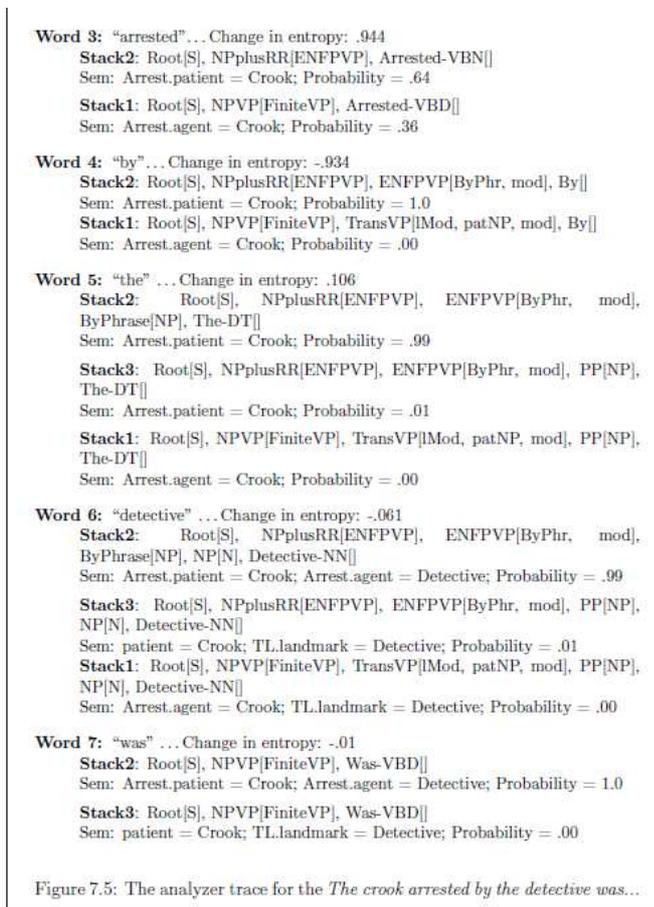
Figura 26 – Processamento *The cop arrested by detective was...*

Fonte: Bryant (2008, p.168)

Em outra análise temos o processamento para *the crook arrested by the detective was...* mostrado na Figura 27. Estruturalmente, as listas na Figura 27 evoluem exatamente da mesma maneira que evoluem na Figura 26. A distribuição de probabilidade, no entanto, é diferente. Em *arrested* há um grande aumento na entropia por causa da incerteza relativa entre a interpretação do verbo principal e a interpretação da relativa reduzida, o que resulta em uma discordância entre sintaxe e semântica a respeito da interpretação apropriada da frase. Posteriormente, não há mais uma discordância entre a sintaxe e a semântica. A palavra *by* torna a interpretação para relativa reduzida adequada tanto sintática quanto semanticamente. A probabilidade da interpretação da relativa reduzida aumenta para 1, e toda a entropia no momento da parada é removida, fazendo com que a diminuição da entropia seja maior. Logo após, o *status quo* é mantido com a exceção da introdução da relativa reduzida com um modificador locativo (Stack3) em *the*. A primeira interpretação da relativa reduzida mantém a probabilidade de 1 quando não se

tem nenhuma mudança notável na entropia. Essas análises mostram como a mudança na entropia após a leitura de cada palavra se correlaciona com mudanças no estado de crença ao longo das análises.

Figura 27 – Processamento de *The crook arrested by detective was...*



Fonte: Bryant (2008, p.169)

Através desse experimento realizado por Bryant (2008), o autor concluiu que as previsões de tempo de leitura fornecidas pelo *Analyzer* são melhores do que as realizadas por McRae et al. (1998), considerando previsões feitas por um modelo probabilístico até o momento. Isso sugere que um modelo baseado em construções pode atuar tanto como um modelo cognitivo de processamento de sentenças e como um modelo psicolinguístico de processamento de sentenças humanas.

Na próxima seção, apresentaremos um outro experimento que também utiliza recursos de uma Gramática de Construções, nesse caso a FCG.

3.2.2 FCG Editor

Nesta seção apresentaremos uma implementação para a análise de construções na *Fluid Construction Grammar* (FCG), a qual já foi apresentada no capítulo 2. Marques e Beuls (2016) apresentam uma proposta para análise e produção de pronomes clíticos no Português Europeu (PE) com base na FCG.

No português europeu (PE), como em outras línguas românicas, clíticos pronominais podem aparecer antes do verbo, próclise, ou depois do verbo, ênclise. No entanto, em outras línguas, o tipo de cliticização é geralmente definido pela finitude do verbo. No PE, a cliticização é mais complexa, pois a mesma forma de verbo pode aparecer anexado pelo mesmo clítico pronominal pré-verbal ou pós-verbal, dependendo do contexto. Isso é ilustrado por Marques e Beuls (2016) no exemplo (16), em que a próclise aparece por causa da conjunção “porque” antes do verbo na segunda oração, não havendo nenhuma pista na primeira levando a uma ênclise.

(16) Eu dei-te este livro, porque tu o querias

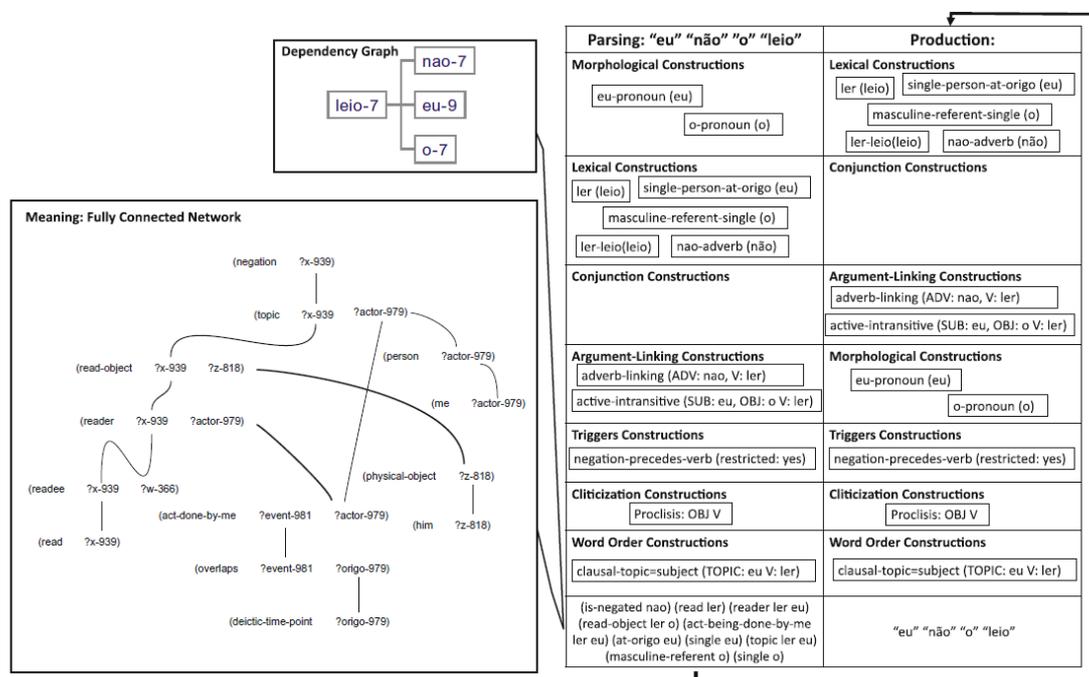
Marques e Beuls (2016) apresentam um modelo computacional inicial, em que o ciclo completo de análise e produção de sentenças é usado como mecanismo de avaliação. A FCG é usada por conta do seu mapeamento parcial entre argumentos sintáticos e semânticos, que, segundo as autoras, aumenta a flexibilidade para lidar com argumentos desconhecidos. A abordagem de Marques e Beuls (2016) se relaciona com analisadores semânticos, na tentativa de recuperar um significado que permite a produção correta da sentença após ser processada. Para esse fim, a gramática inicial é baseada em modelos sintáticos sintagmáticos obtidos a partir de contextos proclíticos encontrados na literatura. O modelo é orientado para a funcionalidade e pode ser facilmente adaptado a um modelo de *crowdsourcing* para expansão automática e não requer anotação de fontes que são propensas a erros, habilitando o sistema a aprender sobre a variabilidade da linguagem com um pequeno número de expressões.

A gramática que as autoras usaram para a análise e produção dos clíticos no PE foi implementada conforme a FCG, uma vez que esse formalismo permitiu representar a informação linguística por meio de estruturas de traços, tanto para unidades básicas, substantivos, ou estruturas mais complexas, orações subordinadas.

A Figura 28 nos mostra como ocorreu a implementação. Existem seis tipos de construções que se aplicam em diferentes ordens de análise e produção. Na análise, as

construções morfológicas são as primeiras a serem aplicadas por correspondência na entrada, os recursos semânticos e sintáticos são adicionados pela construção lexical, e posteriormente podemos observar a realização de uma divisão em duas sentenças pelas construções de conjunção quando uma das palavras é uma conjunção. Na Figura 28 vemos também que as *Argument linking constructions* identificam as conexões semânticas entre os participantes do evento e constroem as dependências gramaticais. Já as *Trigger constructions* encontram acionadores que restringem a posição do clítico por *Cliticization constructions*, enquanto as *word-order constructions* se encarregam das ordenações entre as palavras. Na saída da análise, temos uma rede de significado conectada que é usada para gerar a sentença de volta à produção, juntamente com um gráfico de dependência com ligações entre os dependentes. A produção é semelhante à análise, nela, as construções lexicais são aplicadas primeiro combinando com os significados, e a morfologia das palavras é decidida após o processamento das *argument linking constructions*

Figura 28 – Diagrama com a ordem de aplicação das construções



Marques e Beuls (2016, p. 241)

Ao realizar as análises dos clíticos do PE, Marques e Beuls (2016) observaram os contextos em que ocorre a próclise. Dentre eles, o **Foco frontal**, em que os clíticos

aparecem se o foco vem antes do verbo como em (17) e (18), que diferem apenas no foco. Isso acontece quando a primeira posição da oração (tópico) não é o sujeito. Esse comportamento é alcançado por duas construções: uma que diz que o foco oracional não é o sujeito e adiciona um recurso (restrito +) ao verbo, ativando a próclise; e outra que identifica o foco do predicado, que move o sujeito para o final da sentença. As **Wh Questions**, exemplificadas por (19) e (20) que não são o sujeito e precedem o verbo também ativam automaticamente a próclise, como o caso da *wh-in-situ*.

- (17) Eu dei-te este livro.
- (18) Este livro te dei eu.
- (19) Quem te deu este livro?
- (20) Tu deste-o a quem?

Outros contextos de próclise são de **Negação e advérbios** e **Sujeitos Quantificados**. No primeiro, temos que muitos advérbios podem preceder o verbo (21) ou estarem no final da frase (22). No caso da negação, no PE, ela sempre vem antes do verbo, como em (23). Na implementação de Marques e Beuls (2016), as construções procuram por um advérbio que precede o verbo e verificam se este é um *operator-like* ou um termo de negação, que pode ser usado para restringir o verbo. No segundo contexto, a próclise pode acontecer quando sujeitos quantificados precedem o verbo, como em (24). Assim, a construção procura um quantificador que quantifica um elemento relacionado ao verbo e o precede.

- (21) Eu raramente o leio.
- (22) Eu leio-o raramente.
- (23) Ele não te deu este livro.
- (24) Poucas pessoas o leem.

O último contexto proclítico apresentado por Marques e Beuls (2016) é das **Conjunções subordinadas**, uma vez que conjunções subordinadas também desencadeiam a próclise. Quando uma oração principal é conectada a uma oração subordinada, existem dois verbos diferentes na sentença, dessa forma é possível separar uma oração da outra identificando a conjunção e o verbo em cada uma delas, e introduzindo o traço (restrito +) para o verbo da subordinada, como em (25).

(25) Eu dei-te este livro, porque tu o querias.

Para a análise, Marques e Beuls (2016) testaram o processamento das sentenças e se esse significado foi capaz de produzir a mesma sentença de volta. Diante disso, as autoras optaram por definir as sentenças que poderiam ser produzidas.

A Tabela 1 contém 18 modelos testados pelas autoras. Cada modelo foi avaliado em termos de quantas variações podem se gerar a partir de uma mesma frase (*expected complexity*).

Tabela 1 – Modelos de sentenças analisadas e produzidas

Structure	Examples	Expected Complexity
Noun Phrase		$np = \sum_{i=1}^4 np_i$
NOUN	Livros	$np_1 = n_{noun}$
Art NOUN	O livro	$np_2 = 2 \times n_{nouns}$
DEM NOUN	Este livro	$np_3 = 3 \times n_{nouns}$
NPron	Eu	$np_4 = 8$
Clauses		$n_{clauses} = \sum_{i=1}^{18} cl_i$
1:NP1 V NP	Eu leio livros	$cl_1 = 2 \times np \times n_{cj_vb\wedge pr=np1_pr}$
2:NP1 V CL	Eu leio-o	$cl_2 = np \times n_{cj_vb\wedge pr=np1_pr} \times 11$
3:NP1 V CL NP	Eu leio-te este livro	$cl_3 = np \times n_{cj_vb\wedge pr=np1_pr} \times 6 \times np$
4:NP CL V NP1	Este livro te dei eu	$cl_4 = np \times n_{cj_vb\wedge pr=np1_pr} \times 6 \times np$
5:NP1 ADV CL V	Eu não te vejo	$cl_5 = np \times n_{op_adv} \times n_{cj_vb\wedge pr=np1_pr} \times 11$
6:NP1 ADV CL V NP	Eu não te leio o livro	$cl_6 = 2 \times np \times n_{op_adv} \times n_{cj_vb\wedge pr=np1_pr} \times 6$
7:NP1 ADV V CL	Eu ontem vi-te	$cl_7 = np \times n_{nop_adv} \times n_{cj_vb\wedge pr=np1_pr} \times 11$
8:NP1 ADV V CL NP	Eu ontem li-te este livro	$cl_8 = 2 \times np \times n_{nop_adv} \times n_{cj_vb\wedge pr=np1_pr} \times 6$
9:NP1 V CL ADV	Eu vejo-te raramente	$cl_9 = np \times n_{pos_adv} \times n_{cj_vb\wedge pr=np1_pr} \times 11$
10:NP1 V CL NP ADV	Eu dou-te o livro docemente	$cl_{10} = 2 \times np \times n_{pos_adv} \times n_{cj_vb\wedge pr=np1_pr} \times 6$
11:WH CL V NP	Quem te deu este livro?	$cl_{11} = n_{wh} \times np \times n_{cj_vb\wedge pr=wh_pr} \times 6$
12: WH CL V	Quem o viu?	$cl_{12} = n_{wh} \times n_{cj_vb\wedge pr=wh_pr} \times 11$
13: NP V CL WH	Tu deste-o a quem?	$cl_{13} = np \times n_{wh} \times n_{cj_vb\wedge pr=wh_pr} \times 11$
14: QT NP CL V	Poucas pessoas o lêem	$cl_{14} = n_{dqt} \times np \times n_{cj_vb\wedge pr=np_pr} \times 11$
15: QT NP1 CL V NP	Poucas pessoas te lêem livros	$cl_{15} = n_{dqt} \times 2 \times np \times n_{cj_vb\wedge pr=np1_pr} \times 6$
16: QT NP V CL	Algumas pessoas lêem-no	$cl_{16} = n_{ndqt} \times np \times n_{cj_vb\wedge pr=np_pr} \times 11$
17: QT NP1 V CL NP	Algumas pessoas lêem-te livros	$cl_{17} = n_{ndqt} \times 2 \times np \times n_{cj_vb\wedge pr=np1_pr} \times 6$
Complex Clauses		$n_{complex_clauses} = ccl$
18: Clause CONJ Clause	Eu dei-te o livro, para tu o leres	$ccl = 2 \times n_{clauses} \times n_{conj}$
		$n_{sentences} = n_{complex_clauses} + n_{clauses}$

Marques e Beuls (2016, p. 243)

Como vemos na Tabela 1, os modelos das sentenças que podem ser analisados e produzidos nesse experimento consideram que *n* se refere ao número de cada termo; *cj* *vb* se refere ao tempo verbal conjugado, o *np pr* consiste na pessoa do sintagma nominal, o *n / op adv* o advérbio do tipo negação / operador, e o *n / dqt* refere-se ao quantificador não / descendente)

No modelo, foram considerados dois tipos de artigo (definido / indefinido com 4 formas cada), três demonstrativos, oito pronomes pessoais nominativos, onze clíticos (reflexivos, acusativos e dativos), 6 dos quais são dativos. Com as diferentes ordens que os elementos podem assumir, esperou-se que o sistema aceitasse e gerasse sentenças que poderiam ser obtidas pela introdução automática de construções morfológicas e lexicais. A geração excessiva no modelo foi limitada pela concordância gramatical e pelas redes de significado que direcionavam o processo de produção.

A Tabela 1 ilustra como as sentenças dos contextos proclíticos foram analisadas, na medida que se considerou enquanto estrutura, sintagmas nominais, orações e orações subordinadas.

Em cada contexto proclítico as autoras analisaram como o FCG *Editor*, a partir de uma sentença de entrada, realizou a análise e produziu uma sentença de saída equivalente a primeira. O FCG *Editor* é uma ferramenta em que as análises são realizadas *online* e, por isso, demandaria um grande espaço nessa seção, assim, de igual modo a Marques e Beuls (2016). inserimos aqui o *link* das análises para as sentenças submetidas a esse experimento <https://fcg-net.org/demos/propor-2016/>.

De acordo com as autoras, as análises mostraram que a produção de sentenças proclíticas a partir de uma sentença de entrada ocorreu de forma satisfatória evidenciando que uma Gramática de Construções aplicada a um modelo pode ser útil para tarefas de processamento e produção de sentenças no PE.

A seção seguinte apresentará mais uma implementação, esta, para o reconhecimento de construções. Com base no *Constructicon* da FrameNet Brasil, exibiremos um experimento que considerou algumas Construções de Estrutura Argumental no PB.

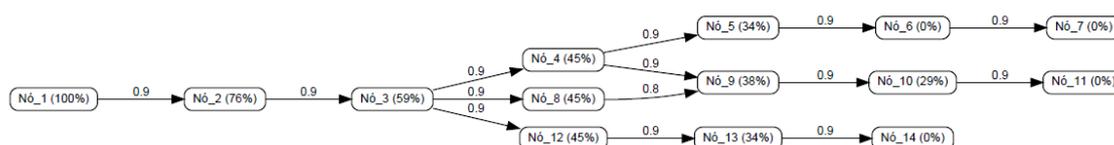
3.2.3 CARMA

O CARMA (Constructional Analyzer using Relations among Multiple AVMs) é um *framework* para a análise de construções gramaticais (MATOS et al., 2017). A partir de um construto, ou seja, uma sentença em língua natural, a sua tarefa é encontrar as construções associadas a esse construto. Com o uso do *Lexicon* e do *Constructicon* da FrameNet Brasil, o CARMA se diferencia de outros *parsers* por utilizar, durante o

processamento, informações semânticas relacionadas aos frames verbais e construcionais dos construtos em análise.

Para a tarefa de reconhecimento de construções por máquina, é utilizado uma técnica de travessia em redes chamada *Spreading Activation*. Essa técnica é aplicada em Redes de Ativação Propagada (RAP), que são geralmente usadas em processos de desambiguação lexical. A Ativação Propagada é um método para pesquisa de redes associativas, redes neurais ou redes semânticas. Nessa técnica, o processo é iniciado pela marcação de um conjunto de nós de origem, por exemplo, os conceitos de uma rede semântica, que possuem um peso e recebem um nível de ativação, em seguida, após a ativação do nó, há a propagação ou espalhamento para outros nós ligados a esse nó de origem.

Figura 29 – Método de Ativação Propagada



Fonte: Matos (2014)

A Figura 29 ilustra o método: a cada passagem por um nó, o nível de ativação é diminuído. Quando o nível de ativação fica abaixo de um certo limiar previamente definido (threshold), ou quando uma condição de parada previamente estabelecida é alcançada, a ativação não é mais realizada. No exemplo hipotético acima, a propagação da ativação é iniciada no nó 1 (100%). Cada link tem o mesmo valor para o peso (0.9), exceto o link entre os nós 8 e 9 (0.8). O fator de decaimento é de 0.85 e o threshold é de 35%. A Figura 29 mostra a ocorrência de 5 ciclos de propagação da ativação. A ativação também pode ocorrer por caminhos alternativos, identificados por meio de marcadores distintos, e só termina quando dois caminhos alternativos alcançam o mesmo nó (MATOS, 2014).

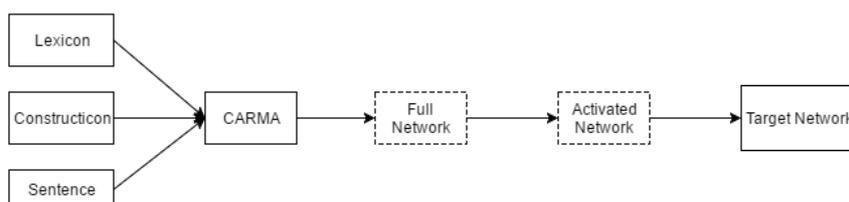
A Ativação Propagada busca modelar a forma como o cérebro funciona, na medida em que impulsos se movem através de toda uma rede de neurônios para recuperar informações específicas. Essa técnica apresenta uma matriz de conceitos dentro da nossa memória como unidades cognitivas, cada uma consistindo de um nó, e os seus elementos ou características associadas, todas ligadas entre si e relacionadas a um conceito original.

Matos (2014), usando a Ativação Propagada em uma proposta de desambiguação lexical baseada na Semântica de Frames, afirma que essa técnica possui um processamento massivo paralelo que ajuda a diferenciá-la de outras técnicas no âmbito da Inteligência Artificial, como também apresenta restrições na comunicação entre as unidades de processamento e mensagens não-simbólicas, tornando o processo independente de uma interpretação específica. Ainda, o resultado obtido no processo pode ser avaliado por um componente externo, um avaliador de caminhos que compara e analisa os caminhos na rede que tenham unidades altamente ativadas. Dessa forma, o uso da técnica da Ativação Propagada para NLU pode ser considerado cognitivamente plausível se os algoritmos e estruturas de dados empregados puderem ser reduzidos a redes conexionistas.

No processo de análise, no CARMA, é criada uma rede semântica cujos nós são as Palavras, Lexemas, Lemas, Unidades Lexicais, Frames, Elementos de Frame, Construções e Elementos de Construção que estão disponíveis na base de dados da FrameNet Brasil. Os links entre os nós são baseados nas relações existentes entre os diversos elementos: relações entre Frames, entre Elementos de Frames, entre Construções e Frames, entre Unidades Lexicais e Frames etc., e um outro link específico entre as palavras é capaz de representar a ordem em que elas aparecem na sentença.

O processamento no CARMA utiliza a técnica da Ativação Propagada, seguindo um procedimento semelhante ao usado para o framework LUDI (MATOS, 2014). O processo é iniciado com a construção da Rede Completa. Esta rede é composta por todos os nós que podem potencialmente ser ativados, a partir das palavras existentes na sentença sob análise. O processo de Ativação Propagada é então aplicado sobre a Rede Completa, gerando a Rede Ativada. Quando o processo de ativação termina, é gerada uma Rede Alvo, a partir das construções e frames que foram ativados por todas as palavras constantes na sentença, conforme delineado na Figura 30 (MATOS et al., 2017).

Figura 30 – Processamento do CARMA



Fonte: Matos et al. (2017)

Na estrutura do CARMA, as construções processadas são formadas por Matrizes de Valores e Atributos (AVMs) cujos signos filhos combinam-se, projetando novas AVMs, as quais contém os valores e atributos das AVMs que se uniram. Essas AVMs correspondem a conjuntos de traços que especificam as características de cada elemento, bem como as possibilidades (ou restrições) de combinações de dois ou mais elementos em uma construção.

As restrições construcionais consideram tanto os aspectos semânticos quanto sintáticos dos signos filhos – ou ECs. Assim, as restrições são tratadas em termo dos ECs que são representados em sintagmas e categorias gramaticais como uma forma de mostrar as generalizações gramaticais de cada construção. Por outro lado, ECs também se encontram unificados a um determinado componente semântico, ou seja, um EF do frame evocado pela construção (MATOS et al., 2017).

Diante disso, podemos explicitar como as relações fundamentais entre as construções – Herança – e de Evocação – entre construções e frames, juntamente com as restrições, servem como evidência positiva para o processamento no CARMA, uma vez que permite maior abrangência dos fenômenos sintático-semânticos.

Almeida (2016) apresenta um experimento em que, a partir da modelagem das construções Transitiva Direta Ativa, Argumento Cindido e Ergativa, um conjunto de construtos licenciados por essas construções foi submetido ao CARMA a fim de testar o seu reconhecimento.

Figura 31 – Exemplos de Construtos

1. A menina quebrou a noz SN _[agente] V SN _[paciente]	Transitiva Direta Ativa
2. O vaso quebrou SN _[paciente] V	Ergativa
3. O estudante quebrou o pé SN _[paciente] V SN _[paciente]	Argumento Cindido

Fonte: Própria autora.

Conforme a Figura 31, a construção Transitiva Direta Ativa pode ser definida [SN [V SN]] em que o primeiro sintagma nominal designa quem praticou uma ação e o segundo sintagma nominal designa quem sofreu a ação como em (1). Já a construção Ergativa tem como característica o fato de o sujeito ser paciente, ou seja, não praticar

Os resultados desse experimento (ALMEIDA, 2016) apontaram para um reconhecimento satisfatório das três construções. Esse reconhecimento foi medido de acordo com o nível de ativação de cada construção. A Tabela 2 mostra o nível de ativação de cada construto e, por conseguinte, a construção que o licencia.

Tabela 2 – Resultados obtidos pelo CARMA no reconhecimento de três construções de estrutura argumental

Construção	Transitiva	Direta	Ativa	ERG	TDA	CAC
Frame evocado: Ação transitiva						
<i>Sentença</i>		<i>Frame evocado pelo verbo</i>				
A faxineira torceu o pano.		Dar forma		-	0,85308	0,83934
A menina arrebentou o lacre.		Danificar		-	0,84366	0,83957
O cachorro machucou o menino.		Causar ferimento		-	0,85082	0,83771
O menino quebrou a noz.		Causar fragmentar		-	0,84067	0,83959
A criança furou os balões.		Danificar		-	0,84059	0,84026
A faxineira trincou o copo.		Danificar		-	0,84059	0,83945
O bandido queimou as provas.		Danificar		-	0,84067	0,84300
A vizinha soltou o cachorro.		Danificar		-	0,84128	0,83813
A professora rasgou o papel.		Danificar		-	0,84059	0,83945
A criança descolou o adesivo.		Danificar		-	0,84129	0,83969
CAC						
Frames evocados: Ser afetado e Parte todo						
<i>Sentença</i>		<i>Frame evocado pelo verbo</i>				
O atleta torceu o tornozelo.		Experenciar ferimento corporal		-	0,85308	0,86228
O menino arrebentou o nariz.		Experenciar ferimento corporal		-	0,84366	0,86238
O jogador machucou o joelho.		Experenciar ferimento corporal		-	0,85082	0,86153
O estudante quebrou o pé.		Experenciar ferimento corporal		-	0,84067	0,86239
O pedreiro furou o dedo.		Experenciar ferimento corporal		-	0,84059	0,86270
O celular trincou a tela.		Tornar-se não-operacional		-	0,84059	0,86200
O computador queimou o HD.		Tornar-se não-operacional		-	0,84214	0,86367
A panela soltou o cabo.		Quebrar		-	0,84128	0,86140
A calça rasgou o bolso.		Tornar-se não-operacional		-	0,84059	0,86083
O sapato descolou a sola.		Quebrar		-	0,84129	0,86076
Construção Ergativa						
Frame evocado: Ser afetado						
<i>Sentença</i>		<i>Frame evocado pelo verbo</i>				
O cabo torceu .		Tomar forma		0,81168	-	-
A corda arrebentou .		Quebrar		0,81193	-	-
O menino machucou .		Experenciar ferimento corporal		0,80990	-	-
O vaso quebrou .		Despedaçar		0,81196	-	-
O balão furou .		Tornar-se não-operacional		0,81270	-	-
A janela trincou .		Despedaçar		0,81180	-	-
O mato queimou .		Pegar fogo		0,81271	-	-
O botão soltou .		Quebrar		0,81036	-	-
O lençol rasgou .		Despedaçar		0,81180	-	-
O cartaz descolou .		Tornar-se solto		0,81208	-	-

Quando observamos o processamento das 30 sentenças, observamos de que maneira a modelagem de construções na FN-Br contribuiu para os resultados obtidos pelo CARMA. Através das duas frentes, Lexicon e Constructicon, integradas, a FN-Br fornece uma base de dados em que os significados são modelados via frames, permitindo maior abrangência dos fenômenos semânticos. Por outro lado, com uma base construcional que conta com um modelo de gramática implementado via sistema de traços por AVMS, é

capaz de proporcionar aos fenômenos linguísticos um enfoque computacional mais eficiente, tanto do ponto de vista sintático, quanto de sua ligação com aspectos semânticos. Dessa forma, os resultados obtidos pelo CARMA se devem ao tipo de tratamento linguístico-computacional fornecido pela FN-Br.

Na próxima seção vamos apresentar outro experimento que também tem o objetivo do reconhecimento de construções para o sueco, e que utiliza os recursos disponíveis no Constructicon sueco.

3.2.4 Swedish Constructicon

Esta seção se propõe apresentar uma pesquisa dentro do projeto *Swedish Constructicon* em que construções são coletadas, analisadas, descritas e publicadas em um recurso disponível gratuitamente.

Segundo Bäckström et al. (2013), a motivação para essa pesquisa surge da ausência de descrição de construções para o sueco. Dessa forma, o experimento inicial apresentado por eles explora o uso de n-gramas híbridos como ferramenta para descoberta de construções em que a busca por candidatas a construção é restrita para padrões parcialmente esquemáticos, ou seja, estruturas em que pelo menos um componente é lexicalmente fixo e pelo menos um componente é esquemático, ou seja, uma categoria morfossintática. Assim, Bäckström et al. (2013) apresentam um experimento projetado para extrair candidatos para uma construção sueca em *corpora* com a utilização de *n-grams* híbridos com o objetivo prático de descobrir construções parcialmente esquemáticas não descritas anteriormente no Constructicon do Sueco.

O Constructicon sueco (SweCxn) (LYNGFELT et al., 2012) é um repositório de construções fundado nos princípios da Gramática de Construção e relacionado à base de dados da Swedish FrameNet++. Segundo Bäckström et al. (2013), uma grande preocupação do SweCxn é considerar os padrões linguísticos que são muito específicos para contar como regras gerais da gramática e gerais demais para serem atribuídos a unidades lexicais individuais. Essas construções são periféricas dentro de uma perspectiva gramatical e lexical e, portanto, são facilmente esquecidas e negligenciadas em gramáticas e recursos lexicais. Atenção especial é dada às construções que representam desafios maiores para a aprendizagem de sueco como L2.

O cenário geral do experimento de Bäckström et al. (2013) é a infraestrutura de recursos do Språkbanken (o *Swedish Language Bank* – banco da língua sueca), um conjunto de recursos e ferramentas na forma de serviços web para acesso, recursos de navegação, edição e anotação automática. As duas facetas da infraestrutura mais relevantes para os presentes propósitos são a infraestrutura de corpus Korp (BORIN et al., 2012b) e a infraestrutura de léxico Karp (BORIN et al., 2012a). Juntos, eles fornecem um conjunto de serviços e recursos interoperáveis para download que permitem que experimentos sejam rapidamente configurados e executados.

A fonte de dados para o experimento é SUC 2.0 (EJERHED e KÄLLGREN, 1997; EJERHED et al., 1992), um *corpus* de texto balanceado para sueco consistindo em 1,17 milhão de tokens que foram anotados manualmente com lemas e MSDs (descrições morfossintáticas). Uma sentença de exemplo no SUC 2.0 é dada na Figura 33 Hur är det då i Mellanöstern? ‘E quanto ao Oriente Médio?’. A primeira parte do MSD é a classe gramatical, por exemplo, VB para är ‘é’. De acordo com Bäckström et al. (2013) o SUC foi selecionado a fim de evitar erros de anotação que confundam os resultados do experimento, mas o experimento pode ser executado em qualquer um dos mais de cem *corpora* de Språkbanken que foram anotados automaticamente com as mesmas informações.

O experimento é baseado no trabalho da StringNet (TSAO e WIBLE, 2009; WIBLE e TSAO, 2010, 2011), em que a noção de *n-grams* híbridos desempenha um papel central. Um *n-gram* híbrido é uma generalização em que não apenas as formas das palavras são incluídas no processo, mas também as informações das camadas de anotação, como mostra a Figura 33 quando se consideram os lemas e as classes gramaticais. No caso deste experimento, então, o *bigram* Hur är (how is) geraria quatro candidatos de construção: *hur vara* (how be), *hur VB* (how VERB), *HA vara* (WH-word be) e *HA VB* (WH-word VERB).

Figura 33 – Anotações SUC 2.0

word	msd	lemma
Hur	HA	hur
är	VB. PRS. AKT	vara
det	PN. NEU. SIN. DEF. SUB+OBJ	den
då	AB	då
i	PP	i
Mellanöstern	PM. NOM	Mellanöstern
?	MAD	

Fonte: Bäckström et al. (2013), p. 4

Uma vez que o objetivo é capturar construções parcialmente esquemáticas, nos experimentos foram descartadas todas as candidatas a construções que são altamente esquemáticas ou altamente lexicais, ou seja, permanecendo *tags* de classe de palavra, por exemplo, *HA VB* ou lemas, *hur vara*. Os autores ainda removeram todos os *n-grams* híbridos que eram subconjuntos de outros *n-grams* híbridos. Um *n-gram* híbrido é considerado um subconjunto de outro se ocorrer como uma subsequência igual ou em itens não conflitantes compartilhando a mesma classe de palavras; por exemplo, *vara_{VB}* é considerado igual a *VB*.

Alguns exemplos de construções candidatas são apresentados na Figura 34. Os *n-grams* híbridos estão ligados à interface Korp para permitir a inspeção de suas instâncias, como vemos na Figura 35, em que temos o número de ocorrências, seguido pela frequência absoluta e a frequência relativa. O resultado completo gerou uma lista de 2500 itens.

Figura 34 – Exemplos de *n-grams* híbridos de SUC 2.0

<i>vara_{VB} ute_{AB} och_{KN} VB</i>	<i>är ute och letar (3)</i>	15	0.93	52.24
<i>vara_{VB} JJ för_{PP} att_{IE}</i>	<i>är viktiga för att (2)</i>	26	1.61	52.83
<i>stänga_{VB} av_{PL} NN</i>	<i>stängt av motorn (1)</i>	11	0.68	52.25

Fonte: Bäckström et al. (2013), p. 5

Figura 35 – As instâncias de *vara*_{VB} *ute*_{AB} *och*_{KN} *VB*

The screenshot shows the ORP interface with a search query: `[lemma contains "vara" & pos = "VB"] [lemma contains "ute" & pos = "AB"] [lemma contains "och" & pos = "KN"] [pos = "VB"]`. The results are displayed in a KWIC (Key-Word In Context) format, showing 15 results. The first result is: `- Svedberg är ute och jagar .` Other results include: `Han är ute och letar efter en tjurkalv s`, `De är ute och samlar ihop namnunders`, `Den som hon bruka ha när hon var ute och högg ven eller hacka på`, `je sommar sprang man omkring här i backarna eller var ute och kajkade med en liten eka`, `Vi måste ju lita på varandra när vi är ute och seglar men viast har vi o`, `"Vi är ute och kontrollerar trafiknykterheten`, `Sextionio och nittiosex är ute och letar ,kom."`, `Men en gång när jag var ute och åkte i vagn hade kuske`, `Eller är de ensamma kvinnor som är ute och raggar sällskap?`, `-Alla är ute och letar efter besparingså`, `förfogar över drygt 3000 fältförsäljare som dagligen är ute och besöker kunder.`, `apporter på lördagseftermiddagen medan de andra var ute och promenerade i den snöfria mell`, `Med glädje noterade jag att hälsohemmen var ute och letade efter själen, myck`, and `Jag är en relativt aktiv motionär, som är ute och springer tre gånger i vecka`. On the right side, the corpus information is shown: `SUC2.0`, `text attributes`, `text: kk82`, `word attributes`, `part-of-speech: verb`, `baseform: vara`, `lemgram: vara (verb)`, `sense: vara`, `initial part: [empty]`, `final part: [empty]`, `dependency relation: ROOT`, and `msd: VB.PRES.AKT`.

Fonte: Bäckström et al. (2013), p. 5

Conforme afirmam Bäckström et al. (2013), a lista de candidatas a construção torna possível percorrer uma grande quantidade de exemplos rapidamente, uma vez que todo *n-gram* híbrido está diretamente vinculado às instâncias do *corpus*. Dos 2500 itens incluídos na lista, 50 construções foram escolhidas como construções candidatas relevantes de acordo com os critérios, ou seja: serem unidades de múltiplas palavras parcialmente esquemáticas e produtivas que são muito gerais para serem atribuídas a um item lexical e muito específicas para serem consideradas como mais esquemáticas.

A lista final de 50 construções relevantes foi extraída em várias etapas. Como o objetivo principal era descobrir construções que são difíceis de encontrar com outros métodos, como afirmam os autores, a identificação de uma candidata a construção também pode colaborar para a descrição de outras construções semelhantes.

No experimento, as instâncias das candidatas a construções exibem propriedades diferentes em relação à função e à forma. Os resultados representam padrões de caráter lexical, idiomático e sintático. Uma evidência de que o método identifica os itens corretos é que algumas das construções de qualificação que já estão descritas no SweCxn foram encontradas. Um desses exemplos é construção de Razão (Rate, no Constructicon de Berkeley), exemplificada em (26) e (27) e formalizada como *RG NN per_{PP} NN*, representando um SN quantificado por um numeral seguido pela preposição sueca *per* e outro SN.

- (26) en gång per dygn
uma vez por dia
- (27) 500 kronor per månad
500 coroas suecas por mês

O experimento também conseguiu identificar construções que não estavam anteriormente descritas no Constructicon sueco, tais como a construção genitiva, formalizada como $RG \text{ år}_{NNGEN} \text{ ålder}_{NN}$, o que representa um numeral seguido da forma genitiva da palavra *år* (ano) e da palavra *ålder* (idade). Essa construção é exemplificada em (28) e (29).

- (28) vid sju års ålder
aos sete anos de idade
- (29) från 17 års ålder
a partir dos 17 anos de idade

O experimento proposto por Bäckström et al. (2013), do ponto de vista metodológico, foi capaz de descobrir algumas construções parcialmente esquemáticas anteriormente não descritas, provando ser uma ferramenta interessante para o trabalho no Constructicon sueco.

Os principais problemas com as candidatas a construção, segundo os autores, são que muitas vezes elas acabam sendo muito sintáticas, por exemplo, uma candidata a construção pode corresponder a um padrão de um sintagma nominal regular, ou muito lexicais, por exemplo, por causa da inflexão interna de uma *multi-word*, ou muito fragmentada, devido à natureza dos *n-grams*. Problema semelhante foi encontrado para o português brasileiro, em um experimento preliminar realizado quando da qualificação desta tese.

O experimento do Constructicon sueco visou antes de tudo identificar possíveis construções que já estavam descritas na ferramenta, como também identificar construções que ainda não haviam recebido um tratamento computacional. Diante dos resultados é possível pontuar que o uso de *n-grams* contribuiu para que diferentes construções fossem identificadas, mas com limitações.

Como temos visto ao longo desta seção alguns trabalhos que visam de alguma maneira utilizar alguma ferramenta computacional que seja capaz de reconhecer diferentes tipos de construções, na próxima seção apresentaremos outro experimento que utiliza um modelo de representação de língua pré-treinado em que é possível submeter diferentes tarefas de identificação de construções.

3.2.5 Mineração de construção

O projeto da FrameNet e do *Constructicon* no alemão tem como objetivo documentar uma gramática baseada em construções do alemão com base em exemplos de *corpus*, incluindo relações entre construções e entre construções e os *frames* por elas evocados. O projeto conta com um banco de dados estruturado e um repositório de construções (*German Constructicon*) e frames (*German FrameNet*) que foi desenvolvido e implementado por Ziem et al. (2019).

O foco do projeto é motivado principalmente pelo objetivo de desenvolver e implementar computacionalmente um fluxo de trabalho constructicográfico, incluindo uma anotação que permite abordar qualquer tipo de construção em vários níveis de esquematicidade, idiomaticidade e abstração. Barteld e Ziem (2021) apresentam uma metodologia orientada para o reconhecimento de construções candidatas para o alemão, denominada *mineração de construções*, que busca identificar construções através de procedimentos que consideram propriedades estatísticas de padrões extraídos de *corpora*

Os autores afirmam que as abordagens de mineração de construção podem ser diferenciadas de duas maneiras: primeiro, em como as candidatas a construção são modeladas e extraídas dos *corpora*, e, segundo, em como elas são classificadas ou selecionadas com base nas estatísticas do *corpus*. Tomando isso como base, os autores apresentam uma estrutura para experimentos com mineração de construção que facilita pesquisas futuras nesta direção. Começam com a introdução das abordagens existentes para o tratamento de construções e para técnicas de mineração e, com base nessas abordagens, apresentam um pacote *python* para tarefas de mineração de construções.

A primeira etapa na mineração de construção é criar uma lista de padrões que modelam a contraparte formal das construções. Esses padrões são criados conforme a extração em um determinado corpus em que, a partir de cada sentença, encontram-se os padrões que se encaixam em um determinado modelo.

Para fins de mineração de construção, Barteld e Ziem (2021) usaram uma metodologia para extrair *n-grams*, que consistem em sequências de *n* elementos consecutivos que aparecem em um texto, em que *n* representa o número definido de elementos. A fim de contabilizar as construções esquemáticas, os autores utilizaram o *StringNet*, um banco de dados de padrões de indexação cruzada extraídos do *British National Corpus*, que faz uso de *n-grams* híbridos, em que os elementos de um *n-gram* são uma forma de palavra, um lema ou uma classe de palavra (POS). Isso permite que o algoritmo extraia diretamente padrões como, por exemplo, *<move> + to + [Verb]*. Os autores afirmam que essas categorias adicionais precisam ser incluídas aos dados no *corpus*, uma vez que *taggers* e analisadores para classes de palavras e informações sintáticas estão disponíveis para muitas línguas, mas isso não se aplica a categorias semânticas.

Após a identificação de cada um dos padrões, o experimento avaliou as construções conforme seu status. O processo de identificação de construções candidatas consistiu em separar padrões significativos, parcialmente ou totalmente esquemáticos, daqueles padrões sem significado ou qualquer função. Essa escolha ocorreu devido à necessidade de se processar uma quantidade razoável de dados reduzindo muitas representações viáveis para um pequeno número de construções reais e produtivas, uma vez que a abundância de padrões extraídos de *corpora* representa um desafio computacional.

Diante dessas colocações, Barteld e Ziem (2021) desenvolveram uma biblioteca de *python* para mineração de construções que permite aplicar abordagens de mineração a um determinado *corpus*. Assim, os usuários podem definir os tipos e parâmetros para filtrar padrões extraídos e chegar a uma lista de construções candidatas geradas a partir de um *corpus*. A ferramenta funciona com *corpora* no formato CoNLL-U (<https://universaldependencies.org/format.html>) em que a entrada consiste na extração de padrões sintáticos e *n-grams* híbridos. Outras categorias como forma da palavra, lema, classes gramaticais e relações de dependência, usadas nos *n-grams* podem ser escolhidas a partir da anotação contida no *corpus*.

Para a extração de construções candidatas, os autores extraíram os padrões de dois *corpora*: a versão da UD (*Universal Dependency*) do Hamburg Dependency Treebank – HDT-UD (Borges Völker et al. 2019) e a Wikipédia do alemão que foi anotada automaticamente de acordo com classe de palavra, lema e relações de dependência usando spaCy (HONNIBAL E MONTANI, 2017).

A Tabela 3 mostra os dez padrões com a melhor classificação de acordo com UIF PMI (FORSBERG et al. 2014), que diz respeito à combinação da frequência de uma instância única (UIF) com PMI, que é uma medida de colocação que modela a força de associação entre os elementos de uma construção, que neste caso é uma generalização multivariada de informação mútua pontual (PMI), resultando em um parâmetro chamado UIF-PMI usado para classificar padrões de acordo com sua relevância e seu potencial construcional. A classificação presente na Tabela 3 considerou os padrões conforme o HDT-UD.

Tabela 3 – Padrões de candidatas a construção *geben*

Pattern	Length	Frequency	UIF	UIF-PMI
geben [Es, obj, PUNCT]	4	25	25	404.57
geben [PRON, obj, PUNCT]	4	32	32	395.02
geben [obl, Auftrag in]	4	19	19	291.32
geben [PROPN, obj, ADP]	4	20	20	229.31
geben [PUNCT, es, obj]	4	14	14	210.38
geben [, PRON, obj]	4	15	15	206.17
geben [PUNCT, PRON, obj]	4	16	16	204.00
geben [, nsubj, ADP]	4	15	15	173.56
geben [PROPN, obj, frei]	4	10	10	169.34
VERB [, PRON, obj]	4	16	16	160.59

Fonte: Barteld e Ziem (2021, p. 11)

Na Tabela 3, os primeiros dois padrões instanciam uma construção candidata que inclui *geben* como a terceira pessoa singular pronominal *es* (it) combinados com a forma finita de *geben* (*es gibt/ há*) que é uma construção com um significado existencial. O terceiro padrão instancia uma construção de verbo “leve” em *Auftrag geben* (contratar). O quarto padrão inclui construções verbais de partículas com *geben*, que compreende instâncias como *abgeben* (entregar algo), *ausgeben* (gastar), *zugeben* (confessar), *freigeben* (limpar) e *aufgeben* (desistir).

Barteld e Ziem (2021) afirmam que esse padrão é altamente polissêmico, uma vez que cada verbo não é semanticamente transparente, por isso, deve ser tratado como uma construção de valência por si só. Os autores observaram que o *slot* do sujeito não é

realizado com uma relação de dependência, ou seja, como função gramatical, mas com a *tag* de classe gramatical para substantivo próprio, indicando que o *slot* do sujeito dessa construção geralmente é preenchido apenas com um nome próprio.

Os padrões seguintes são variações dos primeiros padrões apontando para o uso da construção em orações subordinadas. Finalmente, o nono padrão é um exemplo específico da construção do verbo (*etwas freigeben* (liberar)).

No geral, este experimento mostra como a mineração de construção pode ser usada para reunir construções candidatas, incluindo palavra-alvo, que neste caso consiste no *geben* (dar). Segundo Barteld e Ziem (2021), para transformar essas listas de construções candidatas em um Constructicon é preciso uma constructicografia que considere vários níveis que demandem diversos processos de trabalho.

O experimento de Barteld e Ziem (2021) tentou demonstrar como a mineração de construção é conceituada como um modelo orientado para a identificação de construções candidatas em *corpus*. Inicialmente, essas candidatas a construção assumem a forma de padrões linguísticos com status gramatical indefinido e, estruturalmente, são configurações de *slots* lexicamente preenchidos, isto é, possivelmente esquemáticos de variada complexidade. Especificamente, eles demonstraram que a mineração de construção permite manipular vários parâmetros de constructicografica relevantes, principalmente os tipos de *slots* usados nos padrões. Segundo os autores, o processo de mineração de construção pode ser considerado como ponto de partida e, ao mesmo tempo, a base para um empreendimento constructicográfico com o intuito de desenvolver um repositório estruturado de construções. Eles defendem que a mineração de construção é de fundamental importância porque apenas um processo indutivo permite decidir quais dos padrões detectados são considerados relevantes para um Constructicon, ajudando, assim, a identificar construções candidatas que geralmente são esquecidos e negligenciados

3.2.6 CxGBERT

Nesta seção apresentaremos um experimento que utiliza uma ferramenta denominada BERT (DEVLIN et al., 2019). BERT (*Bidirectional Encoder Representations from Transformers*) é um modelo de representação da linguagem

projetado para pré-treinar representações bidirecionais profundas de texto não rotulados. Como resultado, o modelo BERT pré-treinado pode ser ajustado com apenas uma camada de saída adicional para criar modelos para uma ampla gama de tarefas, como perguntas e respostas e inferências, sem substanciais modificações específicas na arquitetura de tarefas específicas.

Investigar como o BERT funciona, o que capta e a sua capacidade chamou atenção de Madabushi *et al.* (2020). Segundo os autores, embora não seja o primeiro modelo pré-treinado, o BERT utiliza um modelo bidirecional de transformadores que vem fornecendo melhorias significativas no estado da arte de várias tarefas. Apesar da dificuldade inerente em dedicar-se à análise de como funcionam os modelos de *deep learning*, já que grande parte deles gera saídas cujo cálculo não é facilmente verificável, um novo subcampo do PNL, denominado BERTology (ROGERS *et al.*, 2020), surgiu para entender melhor o que o BERT captura a partir da sua estrutura e como é capaz de transferir esse conhecimento para tantas tarefas diferentes por meio de um processo relativamente rápido e eficiente.

O BERT trabalha em duas etapas: Modelo de Linguagem Mascarada (Masked Language Model – MLM), que consiste em um pré-treinamento que envolve “mascarar” aleatoriamente alguns *tokens* estabelecendo o objetivo de prever os tokens originais, e Previsão da Próxima Sentença (Next Sentence Prevision – NSP), que consiste em prever se, dadas duas sentenças, uma é seguida de outra ou não. Essas etapas de treinamento fornecem ao BERT acesso a diversas informações de cunho léxico-semântico. Embora os elementos léxico-semânticos sem dúvida capturem uma grande quantidade de informações linguísticas, eles não capturam todas as informações contidas no texto (GOLDBERG, 1995; FILLMORE, LEE-GOLDMAN E RHOMIEUX, 2012). Esta suposição é central para a abordagem construcionista que não adere a uma divisão estrita entre elementos lexicais e elementos gramaticais, mas, sim, argumenta que eles formam um *continuum* feito de construções, que são pareamentos aprendidos de forma e função ou significado.

No trabalho de Madabushi *et al.* (2020), que utiliza o BERT, encontramos uma técnica chamada *edge probing*, que consiste em um conjunto de tarefas, inspiradas no PNL tradicionalmente estruturado, projetadas para avaliar como os modelos codificam a estrutura da sentença em fenômenos sintáticos, semânticos, locais e de longo alcance. Ao investigar como a adição de informações afeta o BERT e qual é a eficácia dele na identificação de construções, Madabushi *et al.*, (2020) realizaram dois conjuntos de

experimentos. O primeiro conjunto de experimentos visa a adicionar informações construcionais para que o BERT teste o modelo resultante usando tarefas de PLN e o segundo conjunto consiste em sondar modelos do BERT usando uma variedade de métodos para testar quão eficaz é o BERT na identificação de construções.

Os experimentos realizados por Madabushi et al. (2020) classificaram todas as sentenças no corpus *WikiText-103*, em diferentes “documentos”. Esta classificação é baseada em construções, assim, cada sentença é uma instância de uma construção em que os autores usam uma versão modificada do trabalho de Dunn (2017), cujo sistema fornece uma lista de mais 22 mil construções que foram utilizadas para classificar as sentenças. Dessa forma, uma única sentença pode ser classificada como uma instância de várias construções.

Os autores, de modo a processar todas as sentenças do *WikiText-103*, modificaram seu sistema para acelerá-lo, isso resultou em mais de 50.000 sentenças instanciando as construções definidas por Dunn (2017). Nesses experimentos, a coleção de “documentos” consistia em sentenças do *WikiText-103* que são instâncias de uma construção particular: “CxG WikiText”. A Figura 36 fornece uma ilustração de um desses “documentos” com exemplos de sentenças contendo a construção *Personal Pronoun + didn't + V + how* identificada usando uma versão modificada do trabalho de Dunn (2017) com o padrão destacado em negrito.

Figura 36 – Exemplos de sentenças contendo a construção *Personal Pronoun + didn't + V + how*

<p>She didn't understand how I could do so poorly.</p> <p>Kiedis recalled of the situation: “He had such an outpouring of creativity while we were making that album that I think he really didn't know how to live life in tandem with that creativity.”</p> <p>We didn't know how or why.</p> <p>One day she picked up a book and as she opened it, a white child took it away from her, saying she didn't know how to read.</p> <p>In a 1978 interview, Dylan reflected on the period: “I didn't know how to record the way other people were recording, and I didn't want to.</p> <p>And it can be on my album, too, I just didn't realize how it worked. . . At first when I got this, people didn't know that I was an artist, so it was, like, ‘Oh, this songwriter BC.’</p>

Fonte: Madabushi et al. (2020, p. 3)

De acordo com Madabushi et al, (2020) ambos os conjuntos de experimentos foram projetados com base na premissa de que há informações linguísticas significativas codificadas no conhecimento de que duas sentenças são instâncias de uma mesma construção. Os autores afirmam que há evidências teóricas e linguísticas para apoiar essa

afirmação, o que reduz o problema de classificar sentenças em mais de 22 mil construções em uma classificação binária por saber se duas sentenças são ou não instâncias da mesma construção.

Uma vez que o CxGBERT foi treinado em um subconjunto do *WikiText-103*, outra instância do BERT, referida como BERT Clone, foi treinada nos mesmos dados, de modo a ter uma linha de base comparável. Segundo Madabushi et al, (2020) isso foi feito usando os artigos originais da Wikipédia contidos no *WikiText-103*. No entanto, uma vez que algumas sentenças não são instâncias de nenhuma construção, de acordo com o modelo adotado por eles, e, portanto, não incluídas no *CxGWikiText*, estas são removidas do *Corpus*, garantindo que as sentenças seguintes sejam as próximas sentenças instanciadas por construções. Para garantir que qualquer diferença de desempenho entre CxGBERT e BERT Clone seja resultado dos “documentos de construção” e não de alguma estranheza do processo de treinamento, os autores treinaram um terceiro modelo com os mesmos dados que o BERT Clone, mas com as sentenças randomizadas, esse modelo foi chamado BERT Random.

Esses três modelos de BERT foram, portanto, treinados do zero, usando a arquitetura de base do BERT, nas mesmas sentenças. A única diferença é o que constitui um "documento" no pré-treinamento. As sentenças agrupadas com base na Gramática de Construções que são uma instância, no caso da CxGBERT, ou um artigo da Wikipedia, no caso do BERT Clone. Todas as sentenças são repetidas em igual número de vezes no BERT Clone e BERT Random, enquanto no caso de CxGBERT, uma sentença é repetida tantas vezes quanto o número de construções por ela instanciada. Além disso, como o número de sentenças que são instâncias de cada construção varia drasticamente – entre 2 e 50.000 – os autores criaram modelos distintos, separados de acordo com o número de instâncias de cada construção: um conjunto gerado a partir de casos em que a frequência das instâncias encontra-se entre 2 e 10.000 (conjunto inferior) e outro cuja frequência das instâncias está acima de 10.000 (conjunto superior). Conforme Madabushi et al. (2020), o resultado dessa divisão foram seis clones de BERT:

- CxGBERT treinado usando frases que instanciam construções que têm uma frequência de 2 a 10.000 instâncias, chamado de CxGBERT inferior e seus correspondentes BERT Clone e BERT Random inferiores;
- CxGBERT treinado usando frases que instanciam construções que têm uma frequência superior a 10.000 instâncias, CxGBERT superior, acompanhado de BERT Clone superior e BERT Random superior;

Por fim, Madabushi et al, (2020) também realizaram um pré-treinamento da Base BERT, disponibilizado por Devlin et al. (2018), com os dados de treinamento do CxGBERT inferior na tentativa de compensar o conjunto de dados significativamente menor usado no pré-treinamento dos BERT Clones. Para avaliar os oito clones de BERT, Madabushi et al. (2020) usaram as métricas The General Language Understanding Evaluation (GLUE) (Wang et al., 2018) e SQuAD 1.0 (Rajpurkar et al., 2016).

Com o objetivo de avaliar como as informações construcionais alteram o reconhecimento da estrutura da sentença, Madabushi et al. (2020) criaram um conjunto de testes que requerem que o BERT prediga se duas sentenças são instâncias de uma mesma construção. É importante notar que a versão do BERT usada para esses experimentos é a *BERT Base* padrão treinada na Wikipédia e no *Corpus* de “documentos”, não adulterada por qualquer pré-treinamento construcional.

Resultados dos experimentos para medir a capacidade do BERT de distinguir construções e, portanto, reconhecê-las são apresentados primeiramente na Tabela 4, em que temos avaliação dos vários BERT Clones na GLUE e na SQuAD 1.0 junto com sua precisão e perda no Modelo de Linguagem Mascarada (MLM) e na Previsão da Próxima Sentença (NSP). Nessa mesma tabela podemos observar os resultados do BERT nos seguintes *corpora*, *Microsoft Research Paraphrase Corpus* (MRPC), *Corpus of Linguistics Acceptability* (CoLA), *Semantic Text Similarity* (STS), *Recognizing Textual Entailment* (RTE), *School of Science and Technology* (SST)

Tabela 4 – Avaliação BERT

	Lower BERT Base Clone	Lower CxG BERT	Lower BERT Ran- dom	Upper BERT Base Clone	Upper CxG BERT	Upper BERT Ran- dom	BERT Base	BERT Plus CxG 20K	BERT Plus CxG 100K
ML Acc	0.7751	0.7632	0.6957	0.7370	0.6953	0.6805	N/A	0.5971	0.6202
ML Loss	1.0065	1.0638	1.3717	1.1854	1.4242	1.4833	N/A	2.1308	1.9310
NSP Acc	1.0000	1.0000	0.9913	1.0000	0.9938	0.9700	N/A	0.7963	0.8488
NSP Loss	1.24E-5	4.88E-5	1.78E-2	8.08E-6	2.33E-2	9.33E-2	N/A	4.20E-1	3.11E-1
MRPC	83.1210	84.0391	80.9892	83.9428	81.8040	81.4241	87.7551	85.8553	86.0884
CoLA	36.3602	38.3863	30.8298	40.9991	34.3342	37.6885	57.7848	56.7550	59.5925
STS	80.9079	79.8130	28.1752	83.5393	65.2152	29.5565	88.5490	87.4546	86.9574
RTE	60.2888	56.6787	50.1805	64.2599	54.8736	55.5957	62.8159	64.6209	58.8448
SST	91.2844	89.9083	89.9083	91.2844	91.1697	90.5963	91.9725	92.3165	92.4312
SQuAD 1.0	77.8262	76.2176	73.2891	81.9994	75.9331	71.9489	85.3970	84.8543	84.7005

Fonte: Madabushi et al. (2020) p. 8

Os resultados mostram que a forma como as sentenças são agrupadas em documentos tem um impacto significativo, como pode ser visto pelo baixo desempenho de ambos BERT Random. O fato de o BERT Clone superior ter um desempenho melhor do que o BERT Clone inferior pode ser atribuído ao fato de que há mais sentenças nos dados de treinamento superior. Outro fator que pode ser observado é que o CxGBERT falha consideravelmente se comparado ao BERT Clone superior, sugerindo que possivelmente há menos informações contidas nas construções mais frequentes.

A maior lacuna entre o BERT Random e os outros dois modelos está na Similaridade da Semântica do Texto (STS) tarefa que visa testar se duas sentenças têm o mesmo significado, sugerindo que a NSP ajuda a capturar semelhanças semânticas, independentemente de as sentenças serem agrupadas com base na Wikipédia ou em construções.

O desempenho relativamente próximo do CXGBERT e do BERT Clone sugere que a informação de construções e informações de tópicos, extraídas de documentos pertencentes a um único artigo da Wikipédia, parecem ser muito semelhantes ao medido pelo desempenho em tarefas posteriores.

Já na Tabela 5 apresentam-se os resultados do *edge probing* para BERT Clone inferior e CxGBERT inferior. Podemos observar o CxGBERT tendo melhor desempenho em *Named Entity Recognition* (NER), o BERT Clone em *Semantic role labelling* (SRL), os dois com desempenho próximo em *Semantic proto-role* (SPR), e *Non-terminal*, embora o BERT Clone tenha apresentado um desempenho melhor nas outras tarefas.

Tabela 5 – Resultados da *Edge Probing* no BERT Clone inferior e no CxGBERT inferior.

Probe	Lower BERT Base Clone F1	Lower CxGBERT F1	Δ CxG Base	Δ BERT ELMo
PoS	0.807	0.807	0.0	0.0
NER	0.829	0.844	1.5	0.6
SRL	0.687	0.666	-2.1	1.2
SPR	0.786	0.784	-0.2	0.7
Non-terminal	0.609	0.600	-0.9	2.1
NER Task	0.7022	0.7004	-0.18	N/A

Fonte: Madabushi et al. (2020) p. 8

Para testar o desempenho aprimorado do CxGBERT no Reconhecimento de Entidade Nomeada Madabushi et al. (2020) testaram os dois modelos no CoNLL-2003.

Surpreendentemente, o CxGBERT tem um desempenho pior na tarefa, concluindo que a capacidade dos dois modelos de capturar a estrutura da sentença é comparável.

A precisão do BERT Base, sem pré-treinamento adicional, na tarefa de desambiguação de construções é apresentada na Tabela 6. Segundo Madabushi et al. (2020), esses resultados parecem sugerir que o BERT tem uma quantidade surpreendente de informações, sendo capaz de prever se duas sentenças são instâncias da mesma construção com uma precisão de quase 90% após o treinamento em apenas 500 exemplos. O desempenho relativamente baixo da linha de base do Clone parece sugerir que as construções usadas não são muito semanticamente semelhantes, apesar do uso da semântica distributiva para gerá-las. Finalmente, a lacuna no desempenho entre a versão congelada do BERT e as versões inoculadas parece sugerir que o BERT captura informações de construções em sua estrutura interna.

Tabela 6 – Avaliação do BERT Base

	# of CxGs	Full Training	Inoc 5000	Inoc 1000	Inoc 500	Inoc 100	Freeze	No Training	GloVe biLSTM Baseline
2-50	6384	95.0501	92.8571	90.8130	88.9333	72.5564	73.5119	63.9646	69.3163
50-100	2696	93.6573	92.7114	90.4488	88.9466	70.7901	70.8828	63.9651	62.1732
100-1000	8974	94.4451	91.8041	89.4974	86.5612	60.9594	68.2249	58.8478	53.9919
1000-10000	3266	89.7734	87.8292	83.2364	69.7336	57.7771	60.9461	57.7924	54.1405
<10000	21216	94.4075	90.1772	88.6619	85.7325	68.3352	69.3038	61.1708	60.8819
>10000	465	72.5498	72.5100	64.9004	54.9402	53.3865	54.7809	52.8685	52.1306
All	21681	93.4851	89.3409	87.1339	85.8009	64.1137	69.0905	60.1541	62.4242

Fonte: Madabushi et al. (2020, p. 9)

No geral, os experimentos de Madabushi et al. (2020) mostraram que a adição de informações construcionais tem pouco impacto no BERT. No entanto, é interessante observar que a informação linguística disponível para um mecanismo de generalização, como uma rede neural, independentemente de a entrada ser um conjunto de artigos da Wikipédia, como no caso do BERT Clone, ou sentenças instanciando a mesma construção, como no caso do CxGBERT, é muito semelhante. Basta estudar o conjunto diversificado de sentenças que podem ser instâncias da mesma construção para ver que este é um resultado não muito comum.

A segunda questão que Madabushi et al. (2020) pretendiam responder era: Qual a eficácia do BERT na identificação de construções? Os autores descobriram que as construções que são facilmente previstas pelo BERT, usando a tarefa de previsão de

construções, tendem a ter menos instâncias nas sentenças. O experimento revelou que essas construções como *Personal Pronoun + didn't + V + how* também restringem mais o significado da construção.

Por outro lado, aquelas que têm várias sentenças como instâncias, como *Nouns*, que o BERT acha mais difícil de prever, tendem a ser tão gerais que são menos úteis. A altíssima precisão no conjunto de construções que têm menos de 10 mil sentenças como instâncias parece sugerir que o BERT tem uma quantidade substancial de informações pertencentes a construções semanticamente específicas e o fato de que o BERT congelado (*frozen*) funciona mal parece sugerir que esta informação está contida nas camadas internas do BERT e não em sua saída.

Madabushi et al. (2020) afirmam que é impressionante que essas construções possam ser previstas por BERT com 85% de precisão após o treinamento em apenas 500 instâncias. A combinação desses resultados permitiu os autores concluir que o BERT de fato tem acesso a uma significativa quantidade de informações de construções. O impacto desta observação é interessante, pois não apenas mostra as capacidades dos métodos de aprendizagem profunda, mas também mostra que as informações que normalmente são chamadas de informações construcionais podem ser aprendidas a partir da exposição a informações léxico-semânticas.

4 METODOLOGIA

Nesta seção iremos apresentar a metodologia utilizada neste trabalho. Dias da Silva (2006) afirma que é necessário aliar competências de diversas áreas para que a produção efetiva do conhecimento interdisciplinar e a aplicação desse conhecimento no estudo computacional da linguagem possam criar uma estratégia de pesquisa integrada e um sistema computacional apropriado para a compreensão de língua natural. Para esse fim, Dias da Silva (2006) propõe uma estratégia de pesquisa que envolve o equacionamento de três domínios de investigação, a saber: **Linguístico, Linguístico-Computacional e Computacional**.

O Domínio Linguístico envolve a construção de conhecimentos sobre a própria linguagem, é nele que os fatos da língua e do uso são especificados. Nesta fase, a análise dos fenômenos linguísticos é elaborada em termos de modelos e formalismos desenvolvidos no âmbito da teoria linguística. Já o Domínio Linguístico-Computacional envolve questões de seleção e/ou proposição de sistemas formais de representação para os resultados propostos pela fase anterior. Nesta fase, projetam-se as representações linguísticas e extralinguísticas em sistemas formais computacionalmente tratáveis, como o cálculo proposicional e de predicados, as redes semânticas, as regras de reescrita e os *frames*. E por fim, o Domínio Computacional diz respeito à codificação das representações elaboradas durante a fase anterior em termos de linguagens de programação e planejamento global do sistema. Nesta fase, além de transformar as representações da fase anterior em programas computacionais, estudam-se as questões referentes à integração conceitual e física dos vários componentes envolvidos, bem como questões referentes ao ambiente computacional em que o sistema será desenvolvido e implementado.

Um exemplo de como esse equacionamento ocorre pode ser observado na Figura 37, retirada de Dias da Silva (2006).

Figura 37 – Exemplo do tratamento linguístico-computacional de uma configuração sintática

FASE LINGÜÍSTICA	<p align="center">DESCRIÇÃO INFORMAL</p> <p>“Uma frase pode ser composta pela concatenação de um Sintagma Nominal com um Sintagma Verbal. O Sintagma Nominal é o Sujeito da frase. O Sintagma Verbal é o predicado da frase. O Sujeito e o Verbo têm os mesmos traços de Número e Pessoa. O Caso Gramatical do Sujeito é o Nominativo e o Verbo encontra-se na Forma Finita.”</p>
	<p align="center">FORMALISMO LINGÜÍSTICO</p> <p>F → SN SV (↑ SUJEITO) = ↓ ↑ = ↓</p>
FASE LINGÜÍSTICO-COMPUTACIONAL	<p align="center">REPRESENTAÇÃO COMPUTACIONAL</p> <p>Regra Sintática: F → SN SV</p> <p>Especificações: <SN pessoa> = <SV pessoa> <SN número> = <SV número> <SN caso> = nominativo <SV forma verbal> = finita</p>
FASE COMPUTACIONAL	<p align="center">IMPLEMENTAÇÃO NA LINGUAGEM PROLOG¹¹</p> <p>f(P0,P1):-sn(Pessoa,Número,Caso, P0,P1), sv(Pessoa,Número,Caso,P1,P).</p>

Fonte: Dias da Silva, 2006, (p. 126)

Na Figura 37, o ponto de partida é a descrição de um tipo específico de configuração sintática, em que são especificadas linguisticamente as características dessa estrutura. Na sequência, temos a representação dessa configuração através de um formalismo que a torna computacionalmente tratável e, por último, encontramos a implementação, que é realizada através de um programa de computador que a codifica.

Esta pesquisa, que busca apresentar uma metodologia de reconhecimento de construções de estrutura argumental no PB, envolverá as três fases descritas por Dias da Silva para o estudo em Linguística Computacional. Na primeira fase serão apresentadas descrições já realizadas de padrões licenciados pelas construções, levando em consideração aspectos sintáticos, semânticos e pragmáticos. Partindo dessa descrição, apresentaremos, no domínio da representação, uma proposta de formalização de construções no *Constructicon* da FN-Br; e, no domínio da implementação, os resultados do emprego de um algoritmo de Inteligência Artificial para o reconhecimento dessas construções em corpora.

Portanto o que propomos nessa pesquisa, a partir do equacionamento desses três domínios, é que as construções a serem modeladas possam ser inferidas a partir da sua ocorrência em *corpora*.

4.1 FASE LINGUÍSTICA

Esta fase corresponde a análises e descrições que envolvem a explicitação do conhecimento linguístico. Esse conhecimento é descrito através de conceitos, termos, regras, princípios e formalismos. Dotado de uma base gerativista, Dias da Silva (2006) afirma que, no domínio linguístico, a competência e o desempenho devem ser definidos para a construção de sistemas computacionais que os simulem. Neste trabalho, a fase linguística se ancora no paradigma da Gramática de Construções e implica na identificação das construções de estrutura argumental de período simples do Português Brasileiro.

No âmbito da Linguística, o estudo das estruturas argumentais encontrou lugar em diversas abordagens para a sintaxe. A escolha pelas teorias da Gramática de Construções e da Semântica de Frames, para a identificação de construções do ponto de vista linguístico não se deve apenas à sua possível aplicação em outras áreas, mas também por essas teorias permitirem que diferentes estruturas recebam tratamento homólogo.

Goldberg (1995) afirma que uma abordagem construcional pode ser mais eficaz para tratar de estruturas argumentais, pois essas são instâncias de construções que existem independentemente de verbos ou outros itens lexicais predicadores e evocam frames genéricos das ações humanas básicas, possuindo, assim, significado por si mesmas.

Para Fillmore, Kay e O'Connor (1988), os falantes sabem mais sobre uma língua do que as regras bem gerais que dizem respeito a sujeitos, objetos, orações relativas etc. Assim, há muitas expressões que possuem uma forma específica que não pode ser capturada por princípios mais gerais da gramática que não estão vinculados a itens específicos ou significados específicos. Uma abordagem construcional demonstra que tanto estruturas mais idiomáticas como as mais gerais são produtivas e merecem o mesmo tratamento

As construções de estrutura argumental que serão analisadas nesse trabalho já foram descritas anteriormente por estudos em que se consideraram seus aspectos sintáticos, semânticos e pragmáticos.

Como há muitas construções de estrutura argumental no PB, restringiremos nossa análise às construções de período simples da tipologia (S)VO. Partindo das descrições realizadas nessa fase foram elencadas as seguintes construções: Construção Sujeito_Predicado, Construção Transitiva Direta Ativa, Construção Transitiva Indireta, Construção Transitiva Oblíqua, Construção Bitransitiva, Construção Intransitiva, Construção Ergativa, Construção de Objeto Interdito, Argumento Cindido Artefato, Argumento Cindido Parte do Corpo, Argumento Cindido Entidade Atributo, Argumento Cindido Posse Alienável, Construção Resultativa, Construção Predicativa Nominal Artibutiva, Construção Predicativa Nominal Estativa, Predicativa Locativa, Construção Mudança de Estado, Construção Verbo-Nominal Objetiva, Construção Tempo Decorrido, Construção Existencial Apresentacional e Construção Sem Sujeito Intransitiva.

Tais construções, que serão apresentadas no capítulo 5, já foram especificadas de diversas formas na literatura linguística. O nosso intuito ao apresentar essas descrições é embasar, conforme as características de cada construção, a modelagem linguística-computacional realizada na fase posterior.

4.2 FASE LINGUÍSTICO-COMPUTACIONAL

Após o levantamento das descrições das construções de estrutura argumental, a segunda fase do trabalho se concentra no domínio da representação. A fase linguístico-computacional tem o objetivo de transformar a teoria linguística em modelos que, ao mesmo tempo, simulam o comportamento linguístico e seguem as restrições e generalizações previstas pela teoria linguística e pela gramática de uma determinada língua.

De acordo com Dias da Silva (1996), essa fase está entre a representação fornecida pela teoria linguística e o algoritmo e é, além de necessária, estrategicamente positiva, pois, além de funcionar como um projeto arquitetônico que traz as informações necessárias para a construção do sistema, garante maior transportabilidade dos resultados já alcançados para tipos de máquinas diferentes, que possam vir a empregar linguagens de programação também diferentes daquelas originalmente utilizadas nos primeiros testes implementacionais.

Esta pesquisa, que lista entre seus objetivos o de modelar computacionalmente construções de estrutura argumental, se enquadra também na fase representacional,

conforme a descrição dos padrões licenciados pelas construções, levando em consideração aspectos sintáticos, semânticos e pragmáticos. Cabe, então, na fase linguístico-computacional, a construção conceitual do sistema que envolve a seleção ou proposição de sistemas formais de representação para os resultados alcançados na fase anterior.

O sistema formal de representação utilizado nessa fase para este trabalho é o *Constructicon*, a base construcional da FN-Br. No *Constructicon*, foram realizadas duas etapas referentes ao domínio da representação, a saber: modelar as construções de estrutura argumental conforme a descrição já realizada na fase linguística e realizar anotações dessas construções em um *corpus* de treinamento de modo a validar essas construções.

Diversas pesquisas vêm sendo realizadas na FN-Br tanto para modelar construções já descritas na literatura, como as Construções de Argumento Cindido (SAMPAIO, 2010); Objeto Interdito (BROZANTO, 2009); Transitiva, Ergativa, Intransitiva (FERREIRA 2009; CASTILHO, 2010; PERINI 2010), como para identificar construções ainda não descritas, como é o caso da Construção de Mudança de Estado.

A modelagem dessas e de outras construções no *Constructicon* serve a um empreendimento de processamento sintático-semântico de construções no PB. Como o *Constructicon* já foi apresentado no capítulo 2, a modelagem das construções de estrutura argumental será apresentada no capítulo 5, em que se consideraram as características de cada construção conforme a fase linguística.

Outro ponto considerado nessa fase, foi a realização da anotação construcional de um *corpus*. As próximas subseções abordam o corpus utilizado e o processo de anotação construcional.

4.1.1 Corpus Bosque da *Universal Dependencies* (UD) para o português

Esse corpus já anotado manualmente de acordo com as etiquetas da UD conta com 2408 sentenças que posteriormente foram anotadas a partir da modelagem no *Constructicon* na própria ferramenta da FN-Br.

O *corpus* de aprendizagem em que essas construções foram anotadas abarca as duas formas de anotação, a construcional e a anotação da UD. As diferenças entre as formas de anotação ocorrem porque, enquanto o método construcional mapeia uma

construção formalmente, unificando-a com um frame específico que resulta no licenciamento de um construto a anotação da UD é baseada na noção de dependência. Assim a estrutura sintática de uma sentença é constituída por relações assimétricas entre as palavras da sentença, ou seja, temos uma relação de dependência entre um núcleo e um dependente.

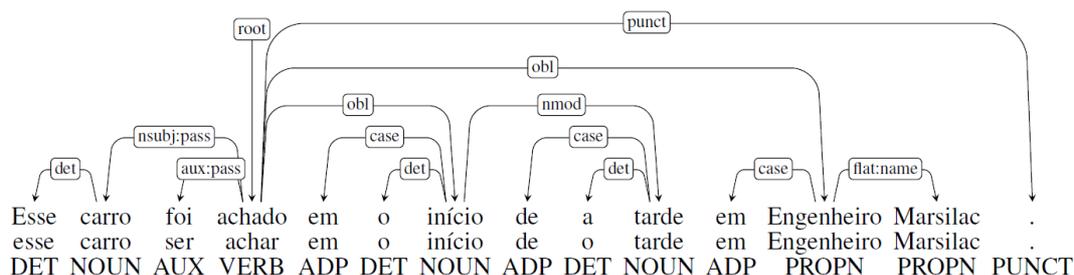
As *Universal Dependencies* “é um projeto que desenvolve anotação *treebank* consistente para várias línguas, com o objetivo de facilitar o desenvolvimento de *parsers* multilíngues, aprendizado translinguístico e pesquisas sobre *parsing* a partir de uma perspectiva de tipologia linguística” (UNIVERSAL DEPENDENCIES PROJECT, 2019).

O *corpus* utilizado nesse experimento segue as diretrizes da UD em um processo que houve a conversão de um *corpus* já existente chamado Bosque em que foi realizado um processamento pelo *parser* PALAVRAS com base na *Constraint Grammar* Rademaker et al. (2017)

O Bosque é um corpus em português, que foi criado e é mantido pela Linguateca, consiste em um subconjunto do Floresta Sintática (AFONSO, et al., 2002) em que temos notícias do Brasil e de Portugal divididas em sentenças sintaticamente analisadas em estruturas arbóreas já anotadas com dependências e manualmente revisada por linguistas.

Rademaker et al. (2017) apresentam esse *corpus*, *UD_Portuguese-Bosque*, em que a partir deste *corpus* convertido, e através da verificação de consistência e visando a total compatibilidade com a especificação da UD, houve uma nova rodada de correções manuais do *treebank* e mudanças nas regras de conversão.

A Figura 38 mostra anotação das sentenças no *UD_Portuguese-Bosque* após o processo de conversão e das alterações realizadas. A sentença “Esse carro foi achado no início da tarde em Engenheiro Marsilac foi anotada com o *parser* PALAVRAS e o projeto UD.

Figura 38 – Anotação no *UD_Portuguese-Bosque*

Fonte: Rademaker et al. (2017, p.200)

4.1.2. Anotação de instâncias do corpus no Constructicon

O *UD_Portuguese-Bosque* foi utilizado para o empreendimento de reconhecimento de construções. Assim, com base na modelagem das construções apresentada no capítulo 5, realizamos a anotação desse *corpus* no *Constructicon* através desse modelo de formalização. Na Figura 39 temos um exemplo de como anotação foi realizada nesse *corpus*.

Figura 39 – Anotação construcional no *UD_Portuguese Bosque*

Save Refresh Hide AS Remove AS Add Cxn ? Label Help									
[156783]	NI	Na época,	111	funcionários	deixaram	a	companhia.		
Transitiva_direta_ativa									
CE					Sujeito			Predicado	
Ação_transitiva.FE					Agente			Paciente	
Sent									

Fonte: Webtool (<http://webtool.framenetbr.uff.br/index.php/webtool/main>, 2020)

A Figura 39 apresenta a anotação do construto *Na época, 111 funcionários deixaram companhia* que corresponde à construção Transitiva Direta Ativa. Observamos que, na primeira camada, se encontram os Elementos de Construção (ECs), anotados respectivamente como Sujeito e Predicado. Na segunda camada, encontramos a relação entre a construção e o *frame* Ação_Transitiva que ela evoca, através da unificação entre o Elemento de Construção e Elemento de *Frame*.

Diante do *corpus* com 2408 sentenças foram realizadas 748 anotações de construções. A metodologia de anotação para as construções no *corpus UD_Portuguese-Bosque* segue as mesmas diretrizes apresentadas no capítulo 2, em que demonstramos

como a anotação de construções é realizada no *Constructicon*, conforme as características sintáticas que foram modeladas para cada construção aliada com as características semânticas através dos *frames*.

4.3 FASE COMPUTACIONAL

A fase computacional corresponde ao domínio da implementação, em que, além de questões sobre a representação por meio de um computador, há também questões que dizem respeito a própria montagem do sistema computacional em que os programas estão inseridos. Para Dias da Silva (2006), é nesta fase que são planejados o sistema e que se codificam as representações elaboradas na fase anterior.

Após todo processo que foi apresentado nas seções anteriores, as construções anotadas no *corpus* foram submetidos a um algoritmo de Inteligência Artificial, que é uma rede neural, para serem classificados conforme as construções existentes no *Constructicon*.

O modelo aplicado a esse trabalho responsável pelo reconhecimento das construções que foram anotadas no *UD_Portuguese-Bosque* é um modelo que, uma vez alimentado por um conjunto de dados, ou seja, o *corpus* de aprendizagem, possa desempenhar a tarefa de reconhecer aquelas construções modeladas no *Constructicon* e aprender esses padrões construcionais.

Ao trabalhar com conjuntos de dados anotados, Pustejovsky e Stubbs (2012) apontam para a existência de três tipos principais de algoritmos de *Machine Learning*: (i) aprendizagem supervisionada, que abrange qualquer técnica que gere um mapeamento de função de entradas para um conjunto fixo de etiquetas que são tipicamente *tags* de metadados fornecidos por humanos que anotam o *corpus* para fins de treinamento; (ii) aprendizagem não supervisionada, que consiste em qualquer técnica que tenta encontrar uma estrutura a partir de um conjunto de entrada de dados não etiquetados e (iii) aprendizagem semi-supervisionada, que baseia-se em qualquer técnica que gere um mapeamento a partir das entradas de dados etiquetados e de dados não etiquetados; uma combinação de aprendizagem supervisionada e não supervisionada.

Em nossa pesquisa, usaremos um algoritmo de aprendizagem semisupervisionada, já, que após a modelagem das construções no *Constructicon*, anotamos o *corpus* de

treinamento dos dados nas bases da FN-Br para que, de posse dele, o algoritmo pudesse reconhecer as instâncias das construções já catalogadas.

A função básica desse algoritmo é a de um classificador, ou seja, ele classifica um padrão como aderente a uma construção existente no *Constructicon* ou não. Como afirmam Pustejovsky e Stubbs (2012), a classificação é a tarefa de identificar a etiquetagem para uma única entidade a partir de um conjunto de dados. Por exemplo, para distinguir spam de não spam em sua caixa de entrada de e-mail, um algoritmo classificador é treinado em um conjunto de dados etiquetados em que os e-mails individuais receberam o rótulo [+ spam] ou [- spam]. É a presença de certas palavras ou sintagmas conhecidos em um e-mail que ajuda a identificar um e-mail como spam. Essas palavras são essencialmente tratadas como recursos que o classificador usa para modelar as instâncias positivas de spam em comparação com as de não-spam.

A partir da modelagem realizada na base construcional da FN-Br, que mapeia as especificidades sintático-semânticas de cada construção, a proposta é que o algoritmo possa classificar os construtos anotados conforme as construções que os licenciam usando, para tanto, também as propriedades semânticas das construções, deriváveis, por exemplo, dos frames evocados pelos verbos que figuram nelas. Por exemplo, a construção Ergativa que licencia “O carro estragou”, evoca o frame de Ser_afetado, o qual é herdado, na rede semântica da FrameNet Brasil, pelo frame de Tornar-se_não-funcional, o qual é evocado por estragar.v. Tal fato, capturado pelo princípio da Coerência Semântica (GOLDBERG, 1995), pode ajudar na determinação de que a probabilidade de “O carro estragou” ser uma instância de construção Ergativa é maior do que aquela de ser uma instância da Construção Intransitiva, por exemplo,, que requer sujeitos agentivos, assim como que os verbos que nelas figuram também o sejam.

O Modelo base (usado em todos os cenários de teste) é constituído de duas partes principais. A primeira é responsável por gerar uma representação da sentença em que o modelo irá identificar os elementos de construção. A segunda usa essa representação para identificar quais *tokens* fazem parte de algum EC. Em todos os casos, a representação gerada pelo primeiro componente do modelo é uma sequência de vetores, cada um deles representando um *token*.

Como esse componente utiliza um modelo BERT, como descrito no capítulo 3, para obter os vetores das palavras, temos uma diferença entre os *tokens* e as palavras da sentença: em “Os meninos correm” as palavras são “Os”, “meninos”, “correm”, mas os *tokens* BERT seriam algo como “O”, “s”, “menino”, “s”, “corre”, “m”. Isso ocorre porque

o BERT usa um tokenizador *WordPiece*¹). Por conta disso, o modelo precisa mapear cada token BERT para um token “palavra”. Esse mapeamento consiste em tirar a média dos vetores do token BERT de cada palavra.

Dessa forma, a média de “menino” + “s”, gera o vetor de “meninos”. Nesse caso, é importante notar que o vetor de “s” não é um vetor do caractere em si, mas um vetor do caractere especificamente na posição em que se encontra na sentença de exemplo. Com essa sequência de vetores de palavras da sentença, o segundo componente do modelo utiliza unidades LSTM (*long-short term memory*) bidirecionais com *hidden_size=20*, junto com uma camada densa, para gerar um vetor de probabilidades para cada palavra. Essas são as probabilidades de que aquela palavra seja parte de cada EC.

Como a modelagem das construções aponta, a especificação dos ECs, que são constituídos por restrições, é primordial para identificar uma determinada construção. Assim, a saída final do modelo é uma sequência que indica, para cada palavra, quais os ECs dos quais ela faz parte, ou nenhum podendo indicar assim qual construção se encontra correspondente naquela sentença.

Esse Modelo Base foi sendo incrementado com outros modelos em que novas informações foram adicionadas, como, por exemplo, as relações da UD e os *frames* evocados pelas construções.

O modelo chamado *UD rel* inclui a informação de relação sintática proveniente das anotações de UD. Para representar as relações, foi utilizado um vetor com 37 valores binários, com cada posição representando uma das possíveis relações e o valor 0 ou 1 indicando se a palavra tem aquela relação ou não.

Isso permite que *tokens* como “do” tenham tanto a relação “case”, quanto “det”, por exemplo. Considerando uma árvore das dependências, essa representação indica apenas o tipo de relação de cada nó com seu nó pai, mas não quem é o pai explicitamente. Em “Mas por não existir um marco legal...”, a palavra “marco” teria a relação “obj”, mas não ficaria explícito para o modelo que “marco” é objeto de “existir”. Esse vetor de 37 posições de cada palavra é concatenado com o vetor de 768 posições obtido usando o BERT.

O modelo *UD order* consiste em reordenar as palavras da sentença antes de serem processadas pelo segundo componente do modelo (bi-LSTM). Para obter os vetores das palavras, a ordem não é alterada, já que é importante para o BERT que os tokens estejam

¹ <https://towardsdatascience.com/wordpiece-subword-based-tokenization-algorithm-1fbd14394ed7>

na ordem correta. A motivação por trás dessa alteração na ordem dos tokens da sentença está justamente em tentar indicar não só o tipo de relação de cada palavra, mas a qual outra ela se relaciona (algo que “UD rel” não faz). Para isso, os tokens são reordenados de maneira que root seja o primeiro, e depois seguido de seus nós filhos e assim até esgotar todos os nós. Essa ordem é basicamente a de uma busca em profundidade na árvore da sentença com base nas relações sintáticas (https://en.wikipedia.org/wiki/Breadth-first_search).

Já o *Token frames Only* é um modelo em que um *frame* é associado a cada palavra da sentença. A identificação do frame evocado por cada palavra é feita automaticamente através da DAISY, algoritmo desenvolvido pela FN-Br (TORRENT ET AL., 2022). Os *frames* são representados da mesma maneira que as relações sintáticas, sendo a única diferença o tamanho do vetor de valores binários.

Por fim, temos o modelo *FrameNet subset* que é uma extensão do anterior, em que, ao invés de associar apenas um *frame* a cada *token*, são associados vários. Esses *frames* são selecionados identificando o menor caminho do frame associado ao *token* até um dos *frames* associados às construções. Caso exista mais de um caminho mínimo, todos serão considerados. Caso nenhum exista, apenas o *frame* do *token* é considerado. Todos os *frames* presentes no(s) caminho(s) são indicados em um vetor similar ao utilizado em *token frames only* e concatenados à representação de cada *token* antes de serem processados pela bi-LSTM.

A Figura 40 mostra exemplos das marcações feitas pelo modelo de reconhecimentos de construções.

Figura 40 – Marcações feitas pelo modelo

Example 50:

- *Original sentence:* O atacante Borges está fora do Campeonato Brasileiro.
- *Construction elements:* O atacante Borges [Predicativa_nominal_estativa.Sujeito] está fora [Predicativa_nominal_estativa.Predicado] do [Predicativa_locativa.Predicado] Campeonato [Predicativa_locativa.Sujeito] Brasileiro [Predicativa_locativa.Predicado].

Example 51:

- *Original sentence:* Como diz o filósofo Giorgio Agamben, a mídia gosta de pessoas indignadas, porém passivas.
- *Construction elements:* Como diz o filósofo Giorgio Agamben, a mídia [Transitiva_obliqua.Sujeito] gosta de pessoas indignadas [Transitiva_obliqua.Predicado] porém passivas.

Fonte: criado pelo modelo

A ideia de um algoritmo capaz de aprender e melhorar com seu próprio desempenho está relacionada a um domínio cada vez mais difundido em ciência da

computação. De acordo com Pustejovsky e Stubbs (2012), aprendizagem de máquina ou *Machine Learning* (ML) é o nome dado a área de Inteligência Artificial relacionada ao desenvolvimento de algoritmos que aprendem ou melhoram seu desempenho a partir da experiência ou de encontros anteriores com dados. Para nossos propósitos, os dados utilizados pelo algoritmo de ML são anotações dos construtos de cada construção de estrutura argumental modelada no *Constructicon*. As anotações fornecidas pelo *Constructicon* e pelo *Lexicon* podem oferecer fontes de dados de entrada ricas para o processo de ML na fase de treinamento, por destacar aspectos específicos de língua natural que são relevantes para a tarefa de aprendizado.

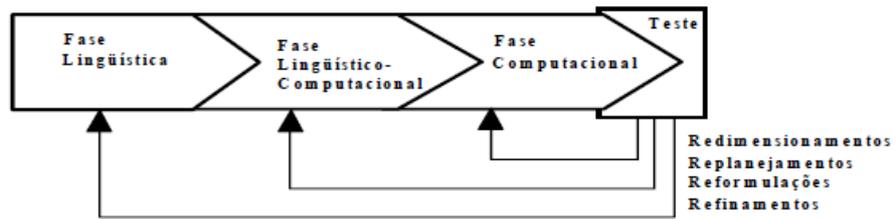
Pustejovsky e Stubbs (2012), ao apresentarem uma metodologia em anotação de corpus para ML, propõem um ciclo em que podemos observar as três fases, linguística, linguístico-computacional e computacional, funcionando ordenadamente. Segundo eles, os recursos utilizados para codificar um fenômeno linguístico específico devem ser ricos o suficiente para capturar o comportamento desejado no algoritmo que estamos treinando.

As descrições linguísticas são tipicamente refinadas a partir da extensa modelagem teórica do fenômeno. Essas descrições, por sua vez, formam a base para os valores de anotação e especificação da linguagem, que são, eles próprios, os recursos usados em um ciclo de desenvolvimento para treinar e testar um algoritmo de identificação ou etiquetagem sobre o texto. Finalmente, com base em uma análise e avaliação do desempenho de um sistema, o modelo do fenômeno pode ser revisado para retreinamento e teste. (PUSTEJOVSKI & STUBBS 2012, p. 24)²

Como podemos observar, as fases linguística, linguístico-computacional e computacional devem ser desenvolvidas sucessiva, progressiva e ciclicamente. Dias da Silva (2006) pontua que as representações parciais resultantes das duas primeiras fases podem ser implementadas na terceira fase e, finalmente, testadas, completando assim um ciclo para que testes de adequação e desempenho sejam realizados com o intuito de aferir o aprimoramento dos resultados alcançados em cada fase. O ciclo proposto por Dias da Silva (2006) pode ser observado na Figura 41.

² These linguistic descriptions are typically distilled from extensive theoretical modeling of the phenomenon. The descriptions in turn form the basis for the annotation values of the specification language, which are themselves the features used in a development cycle for training and testing an identification or labeling algorithm over text. Finally, based on an analysis and evaluation of the performance of a system, the model of the phenomenon may be revised for retraining and testing.

Figura 41 – Fases de construção de um sistema em Processamento de Língua Natural



Fonte: DIAS DA SILVA (2006, p. 125)

Diante da exposição da metodologia desse trabalho, passemos agora para a modelagem das construções no Constructicon que corresponde a fase lingüístico-computacional.

5 CONSTRUÇÕES DE ESTRUTURA ARGUMENTAL SUJEITO_PREDICADO

As construções denominadas de estrutura argumental ganharam destaque principalmente nos trabalhos de Adele Goldberg (1995, 2006). Segundo a autora as construções de estrutura argumental são compostas por seus verbos e seus respectivos argumentos. Como já mencionado no capítulo 2 Goldberg (1995) analisa algumas construções de estrutura argumental com o fim de demonstrar como uma abordagem construcionista pode ser mais eficaz e abrangente para tratar dos fenômenos linguísticos.

A análise de Goldberg contribui para que entendamos que a generalização proposta pela Gramática de Construções nos permite associar uma estrutura sintática a um significado específico, uma vez que as construções de estrutura argumental representam ações humanas básicas e licenciam sentenças independentes de verbos particulares.

Sendo assim, nosso interesse por essas construções ocorre pelo fato dessas representarem estruturas variadas de organização sintática, além disso permitirem generalizações que podem ser descritas em termos de relações de herança e evocação, bem como restrições.

As construções analisadas nesse trabalho são construções de estrutura argumental do Português Brasileiro da tipologia (Sujeito) - Verbo – Objeto ((S)VO) de período simples. Como a nossa proposta visa apresentar um reconhecedor automático de construções, optamos por tratar dessas construções já que essas já foram descritas por diferentes linguistas em diferentes trabalhos, e por serem estruturas representativas de construções esquemáticas.

As análises que serão apresentadas neste capítulo dizem respeito a um conjunto de construções em que primeiramente serão apresentadas a descrição de cada uma delas conforme as suas características sintáticas, semânticas e pragmáticas e posteriormente teremos a modelagem dessas construções no *Constructicon* da FN-Br. No total serão apresentadas 24 (vinte e quatro) construções que correspondem a diferentes estruturas argumentais existentes no PB.

A proposta da modelagem de construções no *Constructicon* da FN-Br possui não só o intuito de oferecer um tratamento linguístico-computacional para as construções do Português Brasileiro, como também utilizar da modelagem realizada para fomentar tarefas de Compreensão de Língua Natural. Dessa forma, nossos objetivos são que o *Constructicon* funcione não apenas como um repositório de construções, mas que sua

estrutura possa se ampliar de modo a abarcar mais construções produtivas na língua que, por mais nucleares e frequentes que sejam, ainda não receberam uma análise computacional, como também permanecer na realização de análises comparativas entre construções de estrutura argumental semântica ou sintaticamente semelhantes, a fim de encontrar as motivações dessas semelhanças.

O modelo do *Constructicon* desenvolvido em Berkeley (FILLMORE ET AL., 2012) como já mencionado nasce da insuficiência do tratamento lexicográfico para determinadas estruturas linguísticas. Sendo assim, o *Constructicon* inicial incorporava apenas aquelas construções que a FN não era capaz de processar através da anotação lexicográfica. Nas palavras de Fillmore et al. (2012), “o *Constructicon* é projetado para lidar com a maioria ou todos os recursos linguísticos relevantes para as construções, já que muitas são idiomáticas ou parcialmente idiomáticas, nem um pouco previsíveis a partir de suas partes componentes.” Assim, a análise realizada no *Constructicon* tinha intuito de trazer informações diferentes acerca das propriedades sintático-semânticas das sentenças.

A partir das discussões feitas em (FILLMORE ET AL., 2012), o *Constructicon* Brasileiro foi implementado, inicialmente, seguindo os mesmos princípios (LAGE, 2013). A FrameNet e o *Constructicon* foram ambos desenvolvidos com a finalidade principal de fomentar pesquisas e tarefas em NLU. Sendo a FrameNet um recurso lexical, suas anotações concernem Unidades Lexicais e os *frames* por elas evocados e, sendo o *Constructicon* um recurso para atender uma série de estruturas linguísticas que não podiam ser incorporadas à anotação lexicográfica, suas anotações concernem estruturas maiores que uma só palavra.

Lage (2013), ao apresentar a implementação do *Constructicon* na FN-Br, propõe alguns critérios para a inserção de estruturas linguísticas nessa ferramenta. De acordo com esses critérios, nem todas as construções deveriam ou poderiam ser modeladas no *Constructicon*, mas apenas aquelas construções em que o padrão de valência das Unidades Lexicais que nelas figuram não fosse suficiente para fornecer todas as informações necessárias para o significado da construção.

No entanto, se considerarmos apenas esses critérios, muitas das realizações das construções que estão contidas hoje no *Constructicon* poderiam ser descritas apenas segundo o padrão de valência dos itens lexicais que funcionam como seus núcleos e, sendo assim, não seriam necessárias na base construcional da FN-Br.

Porém, na perspectiva de Almeida (2016) foi proposto abarcar tanto construções *core* como construções periféricas com o intuito de constituir uma gramática do PB que incluía também padrões construcionais que se manifestam como padrões de valência das Unidades Lexicais. Tal decisão se ancora, ainda, no fato de que os *frames* tipicamente evocados por construções são demasiadamente genéricos para estarem associados a uma UL em particular.

Em seu trabalho, Almeida (2016) apresenta a modelagem de três construções no Constructicon com o objetivo de demonstrar como essa modelagem, a partir das descrições linguísticas existentes, pode auxiliar em tarefas de reconhecimento de construções. Para isso, é necessário não apenas descrever as construções linguisticamente, mas também representá-las, tornando-as computacionalmente tratáveis.

Nas próximas seções, as construções Sujeito_Predicado de período simples serão apresentadas conforme a fase linguística e linguístico-computacional.

5.1 A CONSTRUÇÃO SUJEITO_PREDICADO

Kay e Fillmore (1999) ao analisar as construções nucleadas para o inglês, apresentam a existência de três principais tipos de construções já mencionadas anteriormente. Como as construções analisadas nesse trabalho são herdeiras da construção X_Núcleo, composta por um núcleo que é especificado por um elemento à esquerda, a construção Sujeito_Predicado, que herda a composição binominal abstrata da construção X_Núcleo, fornece um argumento externo, o Sujeito a uma sentença que não o tem, o Predicado.

A construção Sujeito_Predicado sendo altamente abstrata não licenciando nenhum construto específico fornece para as suas construções herdeiras a generalização sintática que será especificada de acordo com os eventos de uso que apresentam essa mesma forma. Como afirmam Fried e Östman (2004) uma vez que, as construções gramaticais sendo reflexos de convenções linguísticas que falantes de uma língua sabem e fazem uso quando se comunicam variam em graus de complexidade e abstração se estruturam por relações de herança que capturam propriedades individuais das construções em comum, as quais são relacionados, mas com padrões gramaticais distintos.

A generalização presente na construção em questão nos permite considerar de forma equânime todas as construções que apresentem a mesma constituição formal, confirmando a proposição de Fillmore (2013) a respeito da necessidade de um modelo de gramática que represente de forma econômica, sistemática e uniforme todos os níveis da estrutura linguística.

A Figura 42 mostra a modelagem da construção Sujeito_Predicado, em que podemos observar seus Elementos de Construção – Sujeito e Predicado – e a relação de herança com outras construções.

Figura 42 – Construção Sujeito_Predicado

Sujeito_predicado []

Definição
A construção de Sujeito_predicado fornece um argumento externo (o **Sujeito**) a uma sentença que não o tem (o **Predicado**).

Exemplo(s)

Elementos da Construção

Predicado [pre] O Predicad é o predicado do **Sujeito**.

Sujeito [sub] O Sujeito é identificado como o argumento externo do Predicado.

Relações

Herda de Argumento_cindido, Bitransitiva, Cláusula_absolutiva, Ergativa, Intransitiva, Mudança_de_estado, Predicativa_locativa, Predicativa_nominal, Transitiva_direta_ativa, por Transitiva_indireta_agentiva, Transitiva_indireta_não-agentiva

Herda de X_núcleo

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Através das relações de herança, apresentaremos como as construções herdeiras da construção Sujeito_Predicado licenciaram os mais diversos construtos no Português Brasileiro da tipologia (S)VO, conforme suas especificidades sintáticas e semânticas.

5.2 CONSTRUÇÃO TRANSITIVA DIRETA ATIVA

A noção tradicional de transitividade verbal nas línguas está relacionada com a presença ou ausência de um objeto enquanto complemento desse verbo. Assim, em uma sentença transitiva direta é obrigatória a existência de um objeto direto.

A construção Transitiva Direta Ativa foi modelada conforme as descrições de Ferreira (2009), Castilho (2010) e Perini (2010), sendo definida pela estrutura [SN[V SN]] em que o sujeito é um agente que pratica uma ação em relação ao paciente. Vejam-se os exemplos (30) e (31).

- (30) Maria quebrou a noz.
 (31) O bandido queimou as provas.

Em construções Transitivas Diretas Ativas, a estrutura argumental dispõe de um predicador, o verbo, cuja valência prevê dois argumentos, sendo o primeiro um SN sujeito e o segundo que não é um sujeito e nem proporcional a um pronome causativo (CASTILHO, 2010). A principal característica dessas construções é o seu caráter causativo, já que o SN-Sujeito é habitualmente agentivo, o causador da ação, e o SN-Objeto Direto é habitualmente o paciente.

Ao propor a modelagem dessa construção, conforme, Figura 43 Almeida (2016) argumenta que consiste em uma construção da família Sujeito_Predicado, nessa medida, a construção Transitiva Direta Ativa traz as informações adicionais de que o elemento antecedente ao Predicado não só é um sujeito, mas que este é agente. Por outro lado, os ECs Sujeito e Predicado da construção Transitiva Direta Ativa são definidos como instâncias licenciadas pelas construções de Sintagma Nominal e Sintagma Verbal com Complemento, respectivamente.

Figura 43 – Construção Transitiva Direta Ativa

Transitiva_direta_ativa [cxn_act_trans_dir]	
Definição	
Tipo de construção [SN V [SN]]. Essa construção exibe um argumento externo Sujeito e um Predicad . O argumento interno é um objeto direto.	
Exemplo(s)	
Elementos da Construção	
Predicado [ce_predicate_act_trans_dir]	O Predicado é composto de um SV e um SN objeto direto.
Sujeito [ce_subject_act_trans_dir]	O sujeito é um SN.
Relações	
Evoca	Ação_transitiva
Herdado por	Objeto_interdito
Herda de	Sujeito_predicado

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

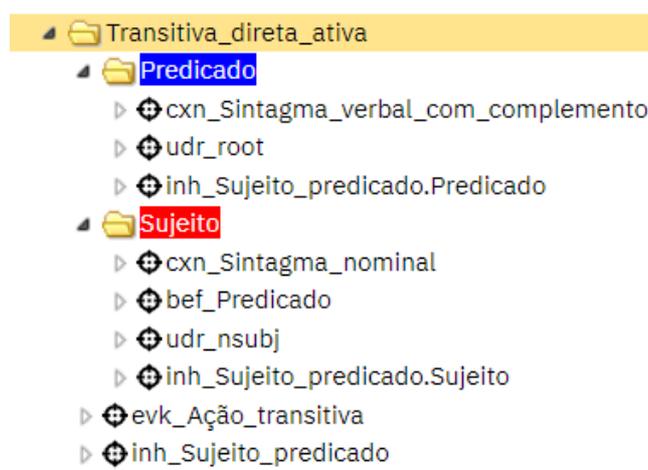
Em relação à contraparte semântica, as construções transitivas, evocam o frame de Ação_transitiva, que é caracterizado por ser um frame não-lexical, ou seja, não evocável por nenhuma UL em particular. Assim, a modelagem nos fornece a unificação em que os sintagmas nominais, Sujeito e Predicado, correspondem aos EFs do frame

Ação_Transitiva, Agente e Paciente. As restrições aplicadas à modelagem da Construção Transitiva Direta e nas outras construções modeladas nesse capítulo foram:

- (i) *CE>Construction*, implementada para determinar os casos em que o signo filho é licenciado por outra construção já definida no *Constructicon*, mostra que os ECs Sujeito e Predicado correspondem às construções Sintagma Nominal e Sintagma Verbal com complemento;
- (ii) *CE_before*, determinando que um dado Elemento da Construção deve vir antes de outro na sentença, indica que o Sujeito deve vir antes do Predicado;
- (iii) *CE>Frame_Family*, que restringe o preenchimento de dado *slot* construcional a todas as ULs que evocam um dado *frame*, bem como a todas aquelas que evocam qualquer um de seus herdeiros, assim, todas as ULs evocadoras de *frame* de Ação_Transitiva ou de seus *frames* filhos podem ocupar esse *slot*.
- (iv) *CE > Inheritance > CE* indica a relação de herança entre construções e a relação de herança entre Elementos de Construção, na construção Transitiva Direta Ativa temos que o *CE_Predicado* é herdeiro do Predicado da Construção Sujeito_Predicado, o mesmo acontece com *CE_Sujeito*;
- (v) *CE > UDRelation* determina qual é a relação sintática do CE de acordo com as *Universal Dependencies*. O CE Predicado recebe a etiqueta *root*, indicando que ele é o núcleo dessa construção e o CE Sujeito recebe a etiqueta *nsubj*, indicando que ele é o sujeito dessa construção.

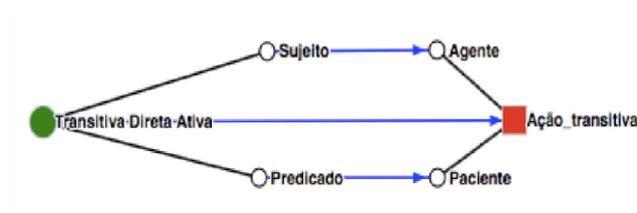
As Figuras 44 e 45 representam a modelagem realizada.

Figura 44 – Restrições aplicadas aos ECs da Construção Transitiva Direta Ativa



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 45 – Frame evocado pela Construção Transitiva



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.3 CONSTRUÇÃO TRANSITIVA INDIRETA

Como uma das funções do *Constructicon* é abranger os mais variados tipos de construções, se fez necessário modelar a construção Transitiva Indireta, uma vez que se mostra produtiva no Português Brasileiro (PB).

A partir das descrições realizadas por Castilho (2010), o qual apresenta estudos sobre a Construção Transitiva Indireta, propomos uma modelagem linguístico-computacional para essa construção. Castilho (2010) define que a construção Transitiva Indireta é composta por um argumento externo sujeito e um argumento interno objeto indireto, o qual seria proporcional a um pronome dativo, como *lhe*. As propriedades sintagmáticas e funcionais dessa construção também são exploradas pelo autor, assim, seria estruturada pela sequência [SN[SV[SP]]] em que o sintagma nominal designa o sujeito e o sintagma preposicionado designa o objeto indireto ou beneficiário, conforme (32) e (33).

- (32) O menino obedeceu à mãe.
 (33) O prefeito respondeu ao jornalista.

A partir desse levantamento, passou-se a uma proposta de modelagem da Construção Transitiva Indireta de acordo com os recursos existentes no *Constructicon* de forma que cada uma das características da construção seja definida em termos de suas partes constituintes, suas propriedades semânticas e relações com outras construções. Sendo assim, o processo de modelagem da Transitiva Indireta iniciou-se a partir da definição das propriedades sintáticas externas destas construções o que demonstrou que tal construção é composta por dois ECs, sendo um o Sujeito e o outro o Predicado. Como mostrado na Figura 46.

Figura 46 – Construção Transitiva Indireta.

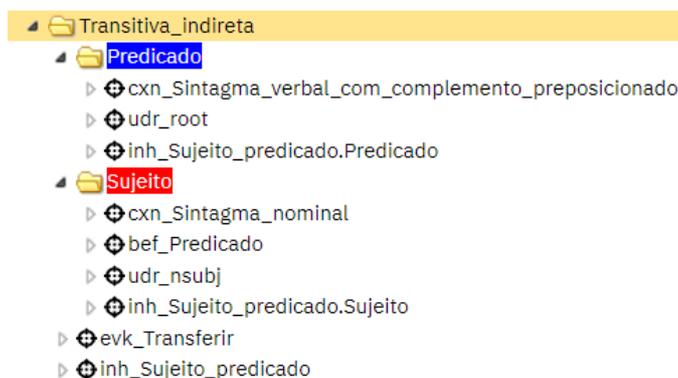
Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Na Figura 46 é possível ver a definição da construção, bem como os ECs que a compõem e suas respectivas definições, além disso, conforme mostrado na aba Relações, a construção Transitiva Indireta é um subtipo da construção mais geral Sujeito_Predicado.

Outro aspecto relevante para o tratamento de construções diz respeito à modelagem de restrições de preenchimento de construções. Assim, é possível estabelecer restrições de constituição para a construção a partir da definição dos ECs, uma vez que a constituição tipifica os signos filhos – os ECs – da construção em termos de outras construções. Dessa forma, enquanto os aspectos formais são considerados para a criação dos ECs, ou signos filhos, a informação semântica é atribuída através da unificação da construção com um *frame*. A aplicação das restrições à construção em questão é mostrada na Figura 47.

Podemos observar na Figura 47 que o EC Predicado constitui uma construção de Sintagma verbal com complemento preposicionado, enquanto, o EC Sujeito é constituído pela construção Sintagma nominal determinado.

Figura 47 – Restrições aplicadas à Construção Transitiva Indireta.



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Para aperfeiçoar a modelagem das descrições apresentadas temos a restrição *CE>Constructicon* que determina quais construções constituem os ECs da Transitiva Indireta. Além da constituição, a unificação, que nos permite relacionar os ECs de uma construção ao *frame* por ela evocado pode ser modelada para que o conjunto de propriedades de uma construção possa apresentar o mesmo tratamento tanto sintático, como semântico. Conforme a modelagem realizada no *Constructicon*, a Construção Transitiva Indireta foi associada ao *frame* Transferir que consiste em um Doador realizando a transferência de um Tema para um Receptor. A Figura 48 mostra o *frame* Transferir.

Figura 48 – Frame Transferir.

Transferir	
Definição	
Este frame envolve um Doador transferindo um Tema para um Receptor .	
Exemplo(s)	
Elementos de Frame Nucleares	
FE Core:	
Doador [donor] excludes: Transferidores	A pessoa que começa com a posse do Tema e faz com que ele esteja na posse do Receptor .
Receptor [recipient] excludes: Transferidores	A entidade que termina na posse do Tema .
Tema [theme] semantic_type: @physical_object	O objeto que muda de dono.
Transferidores [transferors]	O Doador e Receptor concebidos como uma unidade.

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Podemos observar que o *frame* de Transferir apresenta como EFs nucleares, ou seja, aqueles que são obrigatórios para a evocação da cena, pelo menos três participantes, o Doador/Transferidor, Receptor e Tema. A contraparte sintática da Construção Transitiva Indireta que é constituída por dois ECs – Sintagma Nominal e Sintagma Preposicional, os quais precisam ser unificados ao *frame* em questão – será construída a partir da Proeminência (LANGACKER, 1991) de algum elemento em relação a outro. A construção do significado dependerá do nível de especificidade com o qual quisermos retratar a cena. O esquema do *frame* Transferir deve ter todos os EFs nucleares elaborados, mas, de acordo com o que afirma Langacker (1991), uma operação de *construal* permite a elaboração de apenas alguns dos elementos, uma vez que a construção do significado ocorre de acordo com um contexto específico.

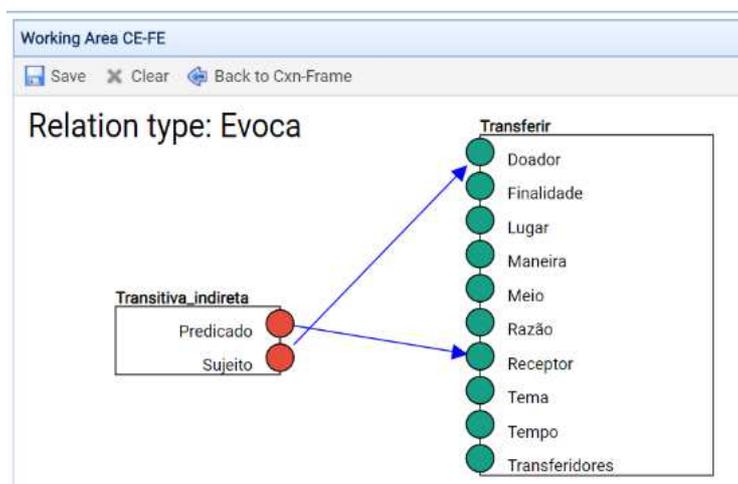
Assim, o EF Tema aparece de forma subfocalizada em função do grau de proeminência do Receptor. O que acontece com semântica da Construção Transitiva Indireta é o que Langacker (1991) chama de *perfilamento*, que consiste no recorte de uma estrutura em uma base conceptual mais ampla, codificando determinados elementos em determinadas situações. Dessa forma, o *frame* Transferir apresenta o perfilamento do Receptor em relação ao Tema, uma vez que este último pode estar incorporado à raiz do verbo, podendo ser identificado através da paráfrase que consiste dos verbos *dar* ou *dever*, seguidos da forma nominal do verbo transitivo indireto como em (34) e (35).

(34) O menino deve obediência à mãe.

(35) O prefeito deu uma resposta ao jornalista.

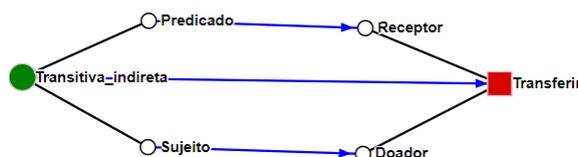
As Figuras 49 e 50 mostram como *frame* Transferir foi relacionado a Construção Transitiva Indireta, dado que o Sujeito que corresponde ao SN foi unificado ao Doador, enquanto o Predicado que consiste do Sintagma Preposicional foi unificado ao Receptor.

Figura 49 – Relação evoca aplicada à Construção Transitiva Indireta.



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 50 – Frame evocado pela Construção Transitiva Indireta.



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.4 CONSTRUÇÃO TRANSITIVA OBLÍQUA

Há uma grande discussão na literatura gramatical a respeito da diferença entre o complemento oblíquo e o objeto indireto. A razão para essa discussão perpassa tanto questões morfosintáticas como questões semânticas. Nesse trabalho, optamos por seguir principalmente as considerações de Castilho (2010), assim como apresentamos a perspectiva diversa defendida por Rocha Lima (2007), que julga essa distinção principalmente pela natureza da preposição que encabeça o Sintagma Preposicionado. Segundo este último autor, o SP será um complemento dativo ou objeto indireto quando esse for por natureza [+humano ou +animado] introduzido pela preposição *a* com papel semântico de *fonte* ou *alvo*. Por outro lado, o sintagma preposicionado será um complemento oblíquo quando introduzido pelas preposições *a*, *de*, *em*, *com* e *para* com papéis semânticos de *tempo*, *locativo* e *companhia*. Para Rocha Lima (2007), portanto, é a preposição funcional que atribuirá o caso dativo ao seu complemento e a preposição lexical, o caso oblíquo.

Já Castilho (2010), mostra a Construção Transitiva Oblíqua como sendo composta por um argumento externo e um argumento oblíquo, que, segundo o autor, é comumente confundido pela Gramática Tradicional como o objeto indireto. A estrutura sintagmática e funcional dessa construção é [SN [SV [SP]]] em que o primeiro SN é o sujeito e o argumento oblíquo corresponde a um sintagma preposicional como em (36) e (37):

(36) Luís gosta de peras.

(37) Pedro precisa de notas.

Nos exemplos (36) e (37) temos os construtos licenciados pela Construção Transitiva Oblíqua. Em (36) o SN *Luís* corresponde ao sujeito, enquanto o SP *de peras* representa o argumento oblíquo, já em (37) o SN *Pedro* constitui o sujeito, enquanto o SN *de notas* constitui o argumento oblíquo. Castilho (2010) argumenta que SNs podem similarmente corresponder a argumentos oblíquos, uma vez que esses também podem apresentar uma característica meio adverbial permitindo às vezes a comutação por um sintagma preposicional complementado por um advérbio dêitico locativo, ou a comutação desse argumento por uma *Preposição + ele/isso/lá* como observado em (38) e (39).

(38) Luís gosta delas.

(39) Pedro precisa disso.

Castilho (2010) chama a atenção para o fato de alguns pesquisadores classificarem o complemento oblíquo ora como adjunto adverbial ora como complemento terminativo. De acordo com Castilho (2010) tal argumento interno não pode ser considerado nem objeto direto nem objeto indireto, por não se possível comutá-lo com os pronomes “o” e “lhes”, argumentando, diferentemente de Rocha Lima (2007), que apenas os sintagmas selecionados pelo verbo serão considerados argumentos. Sendo assim, o argumento oblíquo é uma exigência verbal e não uma atribuição preposicional.

Em Perini (2010), encontramos a *Construção de derrota* e a *Construção paciente com em* que apresentam características semelhantes a Construção Transitiva Oblíqua pelo fato de, sintaticamente, apresentarem argumento interno preposicionado não comutável por “lhe/lhes” como ocorre em (40) e (41).

(40) Daniel apanhou do João.

(41) João bateu no Daniel.

Em (40), temos a Construção de derrota, definida por Perini (2010) como um SN Sujeito que é Paciente + preposição de + SN. Já a Construção paciente com *em* em (41) é definida por um SN-Agente + preposição em + SN. A distinção entre essas duas construções é também semântica, uma vez que, enquanto uma apresenta um sujeito Paciente a outra apresenta um sujeito Agente.

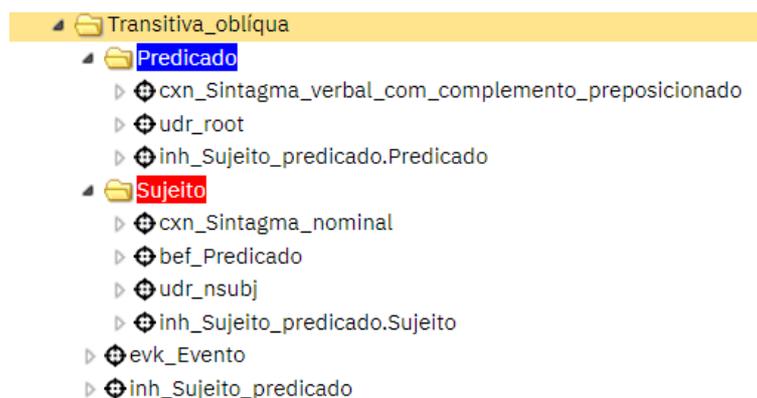
A modelagem da Construção Transitiva Oblíqua no *Constructicon* mapeou que essa construção apresenta um Sujeito que corresponde a um Sintagma nominal e um Predicado que contém um verbo e um Sintagma preposicional. A modelagem mapeou também a relação de Herança com a construção Sujeito_Predicado e a relação Evoca com o *frame* Evento. A Figura 51 mostra a Construção Transitiva Oblíqua com os ECs correspondentes e suas respectivas relações, e a Figura 52 mostra as restrições aplicadas aos ECs da construção.

Figura 51 – Construção Transitiva Oblíqua.

Transitiva_oblua [Transitiva_oblua]	
Definiao	
Tipo de construao [SN[SV[SPrep]]]. Essa construao exibe um argumento externo Sujeito e um argumento interno Predicado. O argumento interno  um objeto oblua.	
Exemplo(s)	
Elementos da Construao	
Predicado [Predicado]	O Predicado  composto de um SV e um SPrep/SN objeto oblua.
Sujeito [Sujeito]	O Sujeito  um SN.
Relaoes	
Evoca	Evento
Herda de	Sujeito_predicado

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 52 – Restrições aplicadas aos ECs da Construção Transitiva Oblíqua.



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.5 CONSTRUÇÃO BITRANSITIVA

A Construção Bitransitiva é também denominada triargumental. Diferente das construções biargumentais, essa construção, segundo Castilho (2010), é organizada de modo que temos um argumento externo e dois argumentos internos que podem ser representados pela estrutura sintática [SN [SV SN[SP]]]. Goldberg (1995), ao fazer uma análise das construções de estrutura argumental do inglês, propõe que a construção bitransitiva apresenta dois objetos e que ambos são SNs. Se considerarmos esse fator, a construção bitransitiva seria pouco produtiva no PB, no entanto, autores como Castilho (2010) e Perini (2010) afirmam a existência da construção em questão, uma vez que, na sua realização, temos a estrutura sintática composta por um Sujeito que corresponde a um Sintagma Nominal, outro Sintagma Nominal correspondente a um Objeto Direto e um Sintagma Preposicionado equivalente a um Objeto Indireto como vemos em (42) e (43).

(42) João deu um livro a Pedro.

(43) O motorista colocou as malas no carro.

Cunha (2017) define a Construção Bitransitiva ou Ditransitiva como uma construção que consiste em um verbo ditransitivo, um argumento agente, um argumento recipiente e um argumento paciente. Essa definição leva em conta o significado da construção, sendo irrelevante a manifestação formal dos argumentos. A autora argumenta que essa definição não corresponde sintaticamente à construção bitransitiva tal como definida por Goldberg (1995). Enquanto no inglês o dativo – recipiente da ação verbal –

é codificado preferencialmente como um SN na posição imediatamente posterior ao verbo e antes do objeto direto, no português do Brasil, o dativo é geralmente codificado antes do objeto direto.

O sentido central da Construção Bitransitiva no português é, como no inglês, aquele em que um agente transfere um paciente para um recipiente. Cunha (2017) afirma que os verbos triargumentais presentes na Construção Bitransitiva são do tipo semântico ação-processo, pois denotam uma ação em que o sujeito animado, intencional causa uma mudança no estado ou na localização do paciente.

Embora o estudo de Cunha (2017) coloque-se como estando alinhado às pesquisas de Goldberg (1995), a autora não pontua como padrão sintático da Construção Bitransitiva a forma em que o objeto indireto representado por um sintagma preposicional se configure na posição posterior ao objeto direto, mesmo esse padrão aparecendo em seus exemplos. A autora defende que o principal padrão sintático para a Construção Bitransitiva no PB seria S V OI OD, em que temos o sujeito o Objeto Indireto e o Objeto Direto, entretanto os estudos como de Castilho (2010) e Perini (2010) apontam outro padrão produtivo dessa construção constituído por S V OD OI.

Como um dos objetivos deste trabalho é apresentar descrições das construções de período simples para o PB, consideramos os apontamentos de Cunha (2017), bem como nos atentamos para as propostas de Castilho (2010) e Perini (2010), uma vez que a distinção dos padrões sintáticos da Construção Bitransitiva pode ser explicada pragmaticamente. Assim, para Cunha (2017), o recipiente é mais proeminente do que no outro padrão.

É importante ressaltar que, no caso da Construção Bitransitiva do português, tomada como objeto de modelagem no *Constructicon* da FrameNet Brasil, o qual é incapaz de representar distinções pragmáticas, tratar como uma mesma construção dois padrões sintáticos diferentes não se apresenta como um problema. Entretanto, essa proposta contraria o Princípio da Não-Sinonímia (GOLDBERG, 1995), assim formulado: se duas construções são sintaticamente distintas, tais construções devem ser também distintas semântica ou pragmaticamente. Goldberg (1995) esclarece que os aspectos pragmáticos das construções envolvem peculiaridades relativas à estrutura da informação, como tópico e foco, e fatores estilísticos, como registro. Até o presente momento, o *Constructicon* é incapaz de modelar tais fatores.

Na modelagem realizada no *Constructicon*, consideramos, portanto, as especificidades sintático-semânticas dessa construção. A composição sintática herdada

da estrutura abstrata da construção Sujeito_Predicado é aquela em que o Sujeito é um SN agentivo e Predicado é um Sintagma verbal bitransitivo, composto por um SN e um SP como podemos observar na Figura 53.

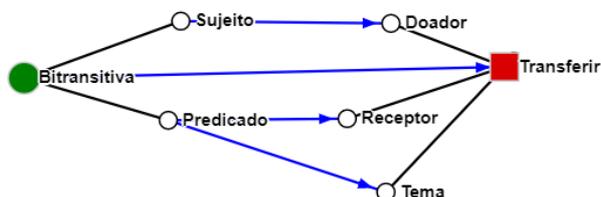
Figura 53 – Construção Bitransitiva.

Bitransitiva [Bitransitiva]	
Definição	Tipo de construção [SN [V SN[SP]]]. Essa construção exibe um argumento externo Sujeito e um Predicado , que é composto por dois argumentos internos.
Exemplo(s)	
Elementos da Construção	<p>Predicado [Predicado] O Predicado por um Sintagma verbal bitransitivo, em que temos um SN e um SP.</p> <p>Sujeito [Sujeito] O Sujeito é um SN agentivo.</p>
Relações	<p>Evoca Transferir</p> <p>Herda de Sujeito_predicado</p>

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

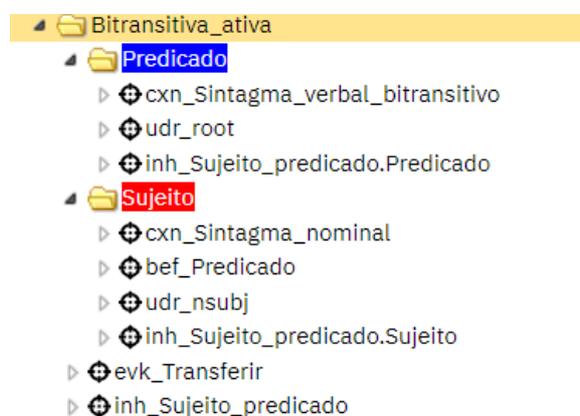
As Figuras Figura 54 e Figura 55 mostram respectivamente a unificação da Construção Bitransitiva ao *frame* por ela evocado e as restrições aplicadas à construção. Como o *frame* evocado pela Construção Bitransitiva é o *frame* Transferir que envolve um Doador transferindo um Tema para um determinado Alvo, a relação Evoca nos permite representar a contraparte semântica da construção em que o SN-Sujeito consiste no Doador e o Predicado, sendo um sintagma verbal bitransitivo composto pelas construções de Sintagma Nominal e Sintagma Preposicional, corresponde os EFs Receptor e Tema do *frame* de que se trata. Como vimos, a Construção Transitiva Indireta também evoca *frame* Transferir. Em relação ao *frame*, a diferença de uma construção para outra está no perfilamento apontado por Langacker (1991): enquanto na Transitiva Indireta o EF Tema é menos proeminente por se encontrar incorporado à raiz verbal, na Bitransitiva o Tema está configurado no SN e apresenta o mesmo grau de proeminência que o Doador e o Receptor, tendo a Construção Bitransitiva os três EFs do *frame* Transferir perfilados.

Figura 54 – Frame evocado pela Construção Bitransitiva.



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 55 – Restrições aplicadas aos ECs da Construção Bitransitiva.



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.6 CONSTRUÇÃO INTRANSITIVA

Enquanto as construções transitivas são biargumentais, em uma construção intransitiva, existe apenas um argumento nuclear, o que explica a razão pela qual a tradição gramatical afirma que nesses casos a ação está integralmente contida na forma verbal, ou seja, a ação não vai além do verbo (CUNHA & CINTRA, 2013).

A construção Intransitiva é caracterizada formalmente pela composição [SN [SV]] que denota atividades que dependem da vontade de um agente como em (44) e (45).

(44) Pedro nadou.

(45) Maria sorriu.

Notamos, tanto em (44) como em (45), que o verbo requer apenas um argumento e seleciona um agente, o qual, conforme Castilho (2010), é o participante desencadeador de alguma ação, capaz de agir com controle. A modelagem dessa construção no *Constructicon* mapeou que uma vez que essa construção não apresenta complemento interno ao verbo, o Elemento da Construção Sujeito é licenciado pela construção Sintagma nominal, enquanto o CE Predicado é, portanto, licenciado pela construção de Sintagma verbal sem complemento, diferentemente do que ocorre com a construção Transitiva Direta Ativa, que tem Predicado licenciado pela construção de Sintagma verbal com complemento. No que diz respeito às restrições, modelou-se o fato de que as restrições CE>Construction e CE>Frame_family, juntamente com a relação de Evocação, definiram que o CE Sujeito consiste de um Sintagma Nominal Determinado e que o CE Predicado, o Sintagma Verbal sem Complemento, é um núcleo verbal, em que o *slot* pode ser preenchido pelas ULs herdeiras do frame Ato_intencional, diferenciando, assim, as construções Intransitivas das Ergativas, conforme Figuras 56, 57 e 58.

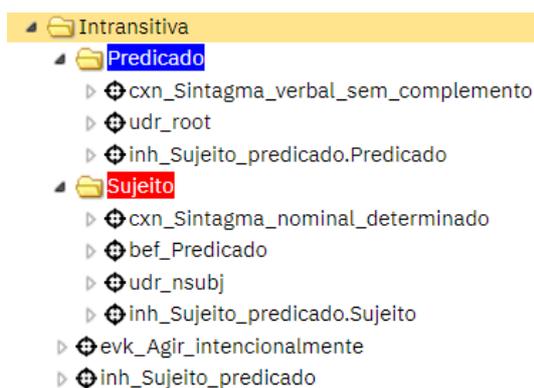
Figura 56 – Construção Intransitiva

Intransitiva [Intransitive]

Definição
Tipo de construção [SN[V]]. Essa construção exibe um argumento externo Sujeito e um Predicado sem complemento interno.
Exemplo(s)
Elementos da Construção
Predicado [Predicate] O Predicado é um Sintagma verbal sem complemento.
Sujeito [Subject] O Sujeito é um Sintagma nominal.
Relações
Evoca Agir_intencionalmente
Herda de Sujeito_predicado

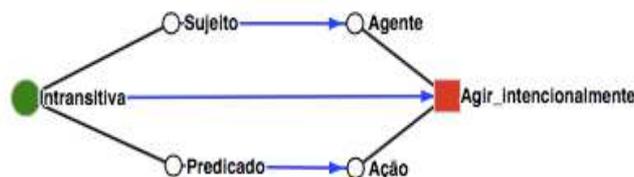
Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 57 – Restrições aplicadas aos ECs da Construção Intransitiva



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 58 – Frame evocado pela Construção Intransitiva



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.7 CONSTRUÇÃO ERGATIVA

A Construção Ergativa é um tipo de construção de estrutura argumental que apresenta a configuração sintática [SN [SV]] em que um sujeito sofre ação de um evento expresso pelo verbo. Diferentemente da construção Intransitiva, que também apresenta uma estrutura monoargumental, o sujeito da Ergativa exhibe comportamento típico do objeto direto dos verbos transitivos. Dessa forma, o argumento interno que é demovido para a posição do sujeito e afetado por uma causa potencialmente externa (PERINI, 2010). A Construção Ergativa licencia construtos como (46) e (47) e, como podemos observar, o sujeito dessa construção não pratica nenhuma ação, sendo paciente.

(46) O vaso quebrou.

(47) O lençol rasgou.

Os verbos usados ergativamente também são considerados inacusativos, uma vez que não há o caráter causativo presente neles. Castilho (2010) afirma que a principal característica das construções ergativas é a não existência de um argumento que controle o estado das coisas, ou seja, não é o vaso quem quebra algo ou o lençol que rasga algo, não estando explícito quem praticou a ação.

No processo de representação dessa construção no Constructicon (ALMEIDA, 2016), conforme Figuras 59 e 60, foi necessário considerar esses aspectos sintáticos-semânticos, uma vez que a construção Ergativa também é herdeira da construção Sujeito-Predicado e evoca o *frame* Ser_afetado, indicando que, enquanto o CE Sujeito

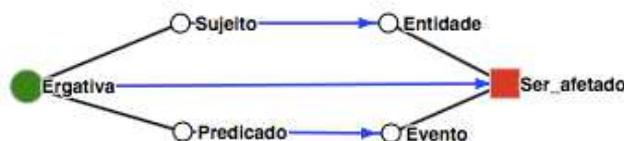
corresponde a entidade que sofre a ação, o CE Predicado corresponde ao evento com apenas um núcleo verbal, que não é manifesto sintaticamente na forma de um novo argumento.

Figura 59 – Construção Ergativa

Ergativa [cxn_ergative]	
Definição	Tipo de construção SN_V. Essa construção exibe um argumento interno Sujeito e um Predicado .
Exemplo(s)	
Elementos da Construção	<p>Predicado [ce_predicate] O Predicado é composto por um SV.</p> <p>Sujeito [ce_subject] O Sujeito é um SN.</p>
Relações	<p>Evoca Ser_afetado</p> <p>Herda de Sujeito_predicado</p>

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 60 – Frame evocado pela Construção Ergativa



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.8 CONSTRUÇÃO DE OBJETO INTERDITO

A Construção de Objeto Interdito (Bronzato, 2009), se relaciona a contextos de detransitivização do predicador verbal, uma vez que ela omite um complemento tipicamente associado ao rompimento das regras de conduta e a tabus sociais. São licenciadas por essa construção sentenças como aquelas em (48), (49) e (50), extraídas do Corpus do Português.

- (48) **Uma senhora que bebe** tanto assim, não é uma senhora.
- (49) Por que **você bebe** tanto, mamãe, você fica cheirando mal.
- (50) **O jovem bebe** para se desinibir e participar dos seus grupos.

As três ocorrências acima apresentam o verbo *beber* relacionado ao rompimento de uma conduta social, associada ao consumo de bebidas alcoólicas, geralmente avaliada de forma negativa pelas pessoas. Segundo Bronzato (2009, p. 78), “as sutilezas semânticas, resultantes da destringitivização de predicadores verbais, são frutos de um processo de mesclagem entre o item lexical e a construção sintática”. Desse modo, a Construção de Objeto Interdito atua sobre os frames evocados pelas ULs verbais presentes nas sentenças licenciadas por ela, no sentido de convidar o interlocutor a uma inferência de que o complemento interdito não é costumeiramente bem-visto.

A construção de Objeto Interdito herda da construção Transitiva Direta Ativa, pois a primeira apresenta a valência básica do verbo alterada, ou seja, verbos como *beber*, *cheirar*, *fumar* são considerados, tradicionalmente, conforme Castilho (2010), biargumentais, isto é, apresentam um argumento externo e um argumento interno. Dessa forma, diferentemente da construção Transitiva Direta Ativa, o objeto na construção de Objeto Interdito é omitido, uma vez que, conforme Bronzato (2009, p. 78), “o apagamento sintático do participante interdito explica-se pela atuação de uma estratégia de proteção de face, representando, pois, uma ação de *polidez positiva*”.

Nesse sentido, a construção evoca o frame de Avaliação_de_moralidade, segundo o qual um Avaliado – Sujeito – é julgado quanto a um Comportamento – Predicado. Em termos de constituição, essa construção é aparentemente similar à construção Intransitiva, uma vez que não apresenta complemento interno ao verbo. Por outro lado, tem o *slot* V preenchido por verbos que preveem, em sua valência, a existência de tal complemento, que foi interdito pela construção.

Assim, o EC Sujeito é licenciado pela construção Sintagma_nominal, enquanto o CE Predicado é, portanto, licenciado pela construção de Sintagma verbal sem complemento, diferentemente do que ocorre com a construção Transitiva Direta Ativa, que tem Predicado licenciado pela construção de Sintagma verbal com complemento.

Até este ponto, em termos de restrições, as construções de Objeto Interdito e Intransitiva são idênticas, o que representa um problema para o modelo. Tal problema é resolvido justamente modelando-se o fato de que, apesar de nuclear um SV sem complemento, o V da construção de Objeto Interdito é, normalmente, biargumental. Na FrameNet, verbos biargumentais evocam frames da família de Ação_transitiva, enquanto verbos monoargumentais, tipicamente presentes na construção Intransitiva, evocam frames herdeiros de Ato_intencional. Assim, as restrições CE>Construction e

CE>Frame_family, juntamente com a relação Evoca, são capazes de diferenciar as construções Transitiva Direta Ativa, Intransitiva e de Objeto Interdito entre si. Esse conjunto de *constraints* foram utilizados como uma estratégia de modelagem de coerção. Uma coerção acontece quando forçamos uma interpretação independente do material linguístico ao qual fomos expostos. Considere que, a princípio, a Construção de Objeto Interdito seria interpretada como um Construção Intransitiva devido a sua estrutura sintática. Porém, o que temos é uma coerção de sentido, gerando uma interpretação diferente das Construções Transitivas e Intransitivas.

De acordo com Michaelis (2002), a coerção não é apenas a resolução de conflitos semânticos, mas a resolução de conflitos entre o significado construcional e o lexical: efeitos de coerção são subprodutos previsíveis dessa combinação. Isso significa que a coerção impõe uma concordância semântica entre a construção e um item lexical, desencadeando uma interpretação forçada, já que, independentemente dos elementos constituintes das estruturas argumentais, essas não são totalmente composicionais e nem totalmente dependentes dos verbos que as instanciam. A maneira pela qual a coerção na Objeto Interdito foi modelada nos permite interpretar, nessa construção, que um verbo prototipicamente transitivo passa a ser usado intransitivamente com a interpretação que o objeto se refere algo que não é aceito moralmente.

Quando o *frame* Avaliação_de_moralidade é mapeado através da relação Evoca e o *frame* Ação_transitiva constitui uma restrição do verbo no EC Predicado e a restrição CE>Inheritance>CE mostra tanto como os CEs Sujeito e Predicado são herdados da Construção Transitiva Direta Ativa, estamos impondo um interpretação distinta daquela que seria se considerássemos apenas as características sintáticas, já que aqui a necessidade do objeto é anulada e o EC Predicado está mapeado para o EF-Comportamento no *frame* Avaliação_de_moralidade, ratificando um dos pilares da Gramática de Construções, o de que o significado da construção é, em parte, independente das palavras que a constituem. Dessa forma, através das restrições apresentadas aqui, foi possível a modelagem adequada da interpretação dessa construção, conforme demonstram as Figuras 61 e 62.

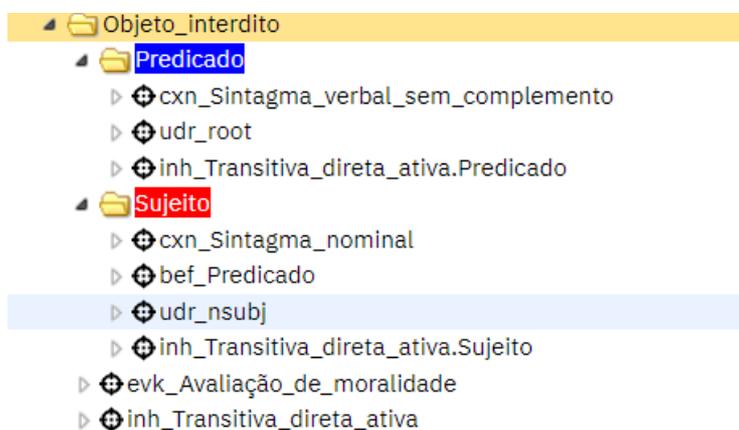
Figura 61 – Construção Objeto Interdito

Objeto_interdito [object_interrupted]	
Definição	Tipo de construção [[SN] [VØ]]. Essa construção exibe a valência alterada dos verbos, apresentando interdição, ou seja, o objeto está implícito na sentença.
Exemplo(s)	
Elementos da Construção	<p>Predicado [Predicate] O predicado é composto por um verbo cujo objeto direto e deliberadamente interditado.</p> <p>Sujeito [Subject] O sujeito é um SN.</p>
Relações	<p>Evoca Avaliação_de_moralidade</p> <p>Herda de Transitiva Direta Ativa</p>

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

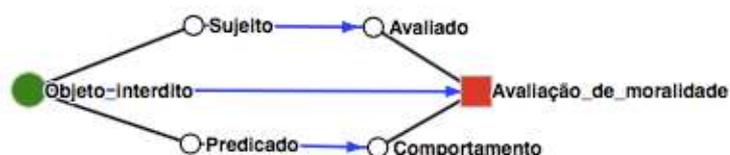
A Figura 61 mostra a modelagem da Construção de Objeto Interdito com a sua descrição e seus respectivos Elementos da Construção e as relações de Evoca e Herança. Por outro lado, nas Figuras 62 e 63 encontramos como as características sintáticas e semânticas foram modeladas através das restrições e da unificação dos ECs com os EFs do *frame* Avaliação_de_moralidade evocado pela construção.

Figura 62 – Restrições aplicadas aos ECs da Construção Objeto Interdito



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 63 – Frame evocado pela Construção Objeto Interdito



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.9 CONSTRUÇÃO DE ARGUMENTO CINDIDO

As Construções de Argumento Cindido (CACs) são definidas por Sampaio (2010) como construções que apresentam dois argumentos sintáticos e apenas um argumento semântico com função de paciente. Apesar de essas construções apresentarem uma sintaxe semelhante com a construção Transitiva, não existe nenhuma ação praticada pelo sujeito que afete um paciente, o que ocorre na verdade, é que uma mesma entidade se comporta como sujeito e paciente como ilustram os exemplos (51) e (52).

(51) O celular quebrou a tela.

(52) A calça rasgou o bolso.

Sendo assim, o que caracteriza as CACs, além da sua estrutura biargumental e a expressão de um evento que afeta um paciente, evocando assim o *frame* Ser_afetado

Em sua análise, Sampaio (2010) apresentou a postulação de uma família de CACs, que levou em consideração, principalmente, as classes dos verbos licenciados, o tipo semântico da entidade afetada e os padrões de uso das diferentes versões da CAC. Os dados obtidos permitiram a identificação e caracterização de quatro membros dessa família de construções: a CAC-artefato; a CAC-parte do corpo; a CAC-posse alienável, e a CAC-entidade/atributo.

As quatro construções dessa família apesar da semelhança possuem características semânticas distintas, tais construções apresentam a estrutura sintática [SN [V [SN]]], mas evocam *frames* diferentes. Nesse sentido, para modelar essas construções no *Constructicon* criamos primeiro uma CAC genérica responsável por fornecer a estrutura para as outras CACs conforme a Figura 64.

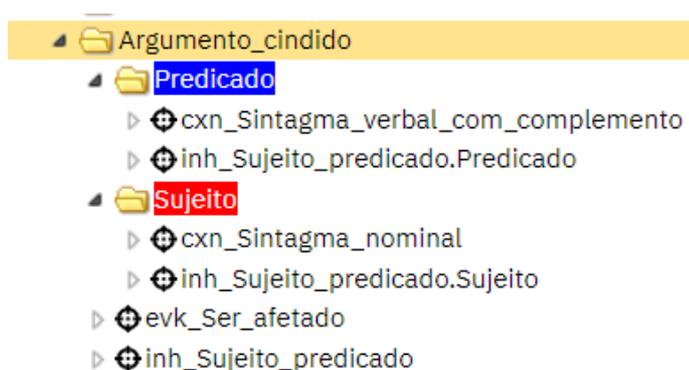
Figura 64 – CAC genérica

Argumento_cindido [Argumento_cindido] [131164]	
Definição	
Tipo de construção abstrata composta por um SN Sujeito e um Predicado composto por um Verbo e um SN.	
Exemplo(s)	
Elementos da Construção	
Predicado [Predicado]	O Predicado é composto de um SV e um SN
Sujeito [Sujeito]	O sujeito é um SN.
Relações	
Evoca	Ser_afetado
Herdado por	Argumento_cindido_artefato, Argumento_cindido_entidade_tributo, Argumento_cindido_posse_alienável
Herda de	Sujeito_predicado

Fonte: Webtool (<http://webtool.framenetbr.uff.br/index.php/webtool/main>, 2020)

Da mesma forma que a Construção de Argumento Cindido mais genérica fornece material sintático para as CACs específicas por herança completa (Kay e Fillmore, 1999), essa relação também permite que o *frame* Ser_ afetado também seja herdado pelas outras construções. A Figura 65 apresenta as restrições aplicadas à CAC genérica em que podemos observar a relação da construção com o *frame* evocado. Assim, as outras Construções de Argumento Cindido serão especificados conforme o *frame* particular de cada construção, uma vez que a informação do frame Ser_ afetado já foi modelada na construção mãe.

Figura 65 – Restrições aplicadas à CAC genérica



Fonte: Webtool (<http://webtool.framenetbr.uff.br/index.php/webtool/main>, 2020)

A CAC-artefato, como apresentada nos exemplos (51) e (52), tem como característica o fato de o dano ser conceptualizado pelo usuário como uma espécie de evento espontâneo, e não como resultado da ação de um agente. Em relação aos verbos, esses constituem um grupo bastante definido de verbos que evocam *frames* relacionados à noção de dano. No caso específico da CAC-artefato, a construção como as outras dessa

família evoca um *frame* próprio além do *frame* Ser_afetado. Tanto na CAC como na Ergativa Canônica temos esse mesmo *frame* sendo evocado. Esse *frame*, por sua vez, possui apenas um EF obrigatoriamente expresso: a entidade afetada. Outra característica primordial da CAC-artefato é a de evocar também o *frame* Parte_Todo, em que o Sujeito está associado a um artefato e uma parte específica desse artefato se encontra expresso no objeto do verbo, estando associado ao EF Parte do *frame* em questão.

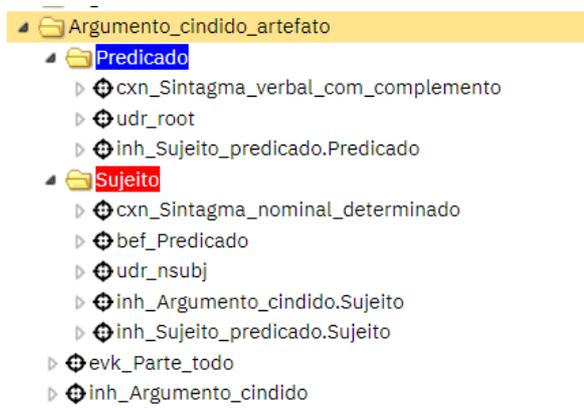
Figura 66 – Construção Argumento Cindido-artefato

Argumento_cindido_artefato [cxn_split_object] [17678]	
Definição	Tipo de construção de Argumento_cindido [SN[V SN]]. Essa construção exibe um argumento externo Sujeito e um Predicado . O argumento interno é um objeto direto. A construção apresenta uma relação PARTE/TUDO entre os argumentos.
Exemplo(s)	
Elementos da Construção	<p>Predicado [ce_predicate_split_object] O Predicado é composto de um SV e um SN objeto direto.</p> <p>Sujeito [ce_subject_split_object] O sujeito é um SN.</p>
Relações	<p>Evoca Parte_todo</p> <p>Herda de Argumento_cindido</p>

Fonte: Webtool (<http://webtool.framenetbr.uff.br/index.php/webtool/main>, 2020)

A Figura 66 mostra a Construção Argumento Cindido-artefato conforme suas características sintáticas e semânticas. Na Figura 67, podemos observar as restrições que foram aplicadas a CAC-artefato, indicando inclusive o *frame* Parte_Todo responsável pela interpretação correta dessa construção.

Figura 67 – Restrições aplicadas aos ECs da CAC-artefato



Fonte: Webtool (<http://webtool.framenetbr.uff.br/index.php/webtool/main>, 2020)

A CAC-parte do corpo pode ser caracterizada como uma Construção de Argumento Cindido, com verbo que evoca o *frame* de dano sofrido em uma parte do corpo. Seu sujeito é um ser vivo e seu objeto uma parte específica do corpo desse ser, configurando uma relação de posse inalienável, como ocorre em (53) e (54).

(53) O atleta torceu o tornozelo.

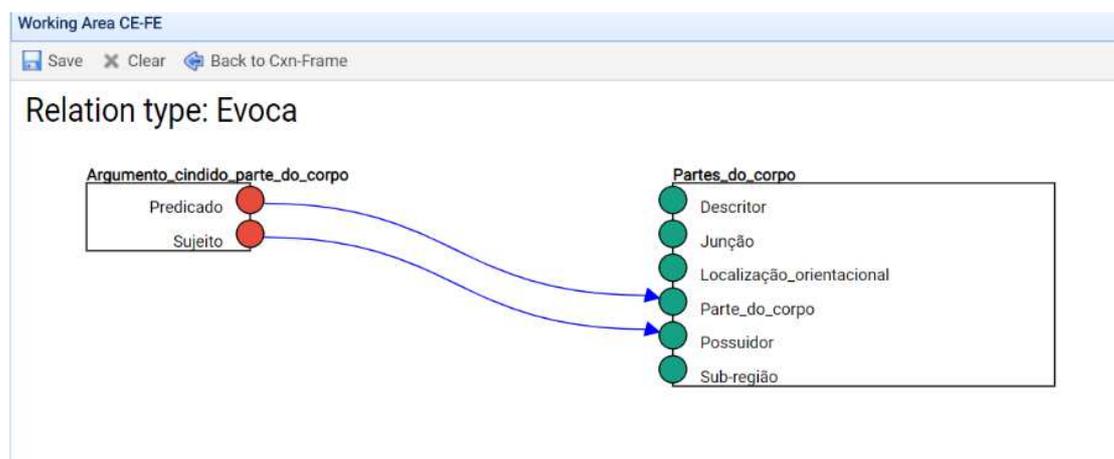
(54) A cozinheira cortou o dedo.

Do mesmo modo que ocorre com a CAC-artefato, a CAC-parte do corpo apresenta a mesma configuração semântica. Como demonstram os exemplos (53) e (54), por mais que encontremos neles um sujeito que pode ser tipicamente agentivo, a causatividade não está presente nessa construção. Em outras palavras, por mais que seja possível que o atleta e a cozinheira tenham machucado a si mesmos, a leitura que fazemos desses exemplos é que um evento externo ocorreu e levou ao dano na parte do corpo de alguém.

Sampaio (2010) afirma que caso o sentido pretendido seja causativo, o usuário precisa indicar isso de alguma forma, marcando o verbo uma ação intencional “Jorge quebrou o próprio braço”, ou explicitando a dissociação semântica entre sujeito e objeto “Jorge quebrou o braço de Vicente”. Em relação ao *frame* da construção, a CAC-parte do corpo evoca o *frame* Parte_do_corpo que é caracterizado pela posse de uma determinada parte do corpo por um ser prototipicamente agentivo. Assim, o EF Parte do corpo é mapeado para a parte do corpo afetada e o EF Possuidor é mapeado para o ser a quem a parte do corpo pertence.

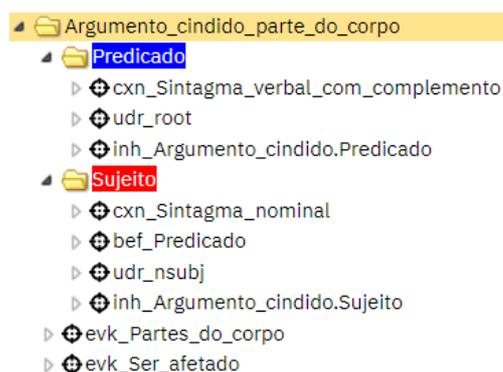
Na Figura 68 vemos como o *frame* Partes_do_corpo foi unificado aos ECs da CAC-parte do corpo, diferenciando-a assim, das demais CACs, e na Figura 68 temos as restrições que foram aplicadas aos ECs dessa construção.

Figura 68 – Relação Evoca CAC-parte do corpo



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 69 – Restrições aplicadas aos ECs da CAC-parte do corpo



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

O terceiro subtipo de CAC apresentada por Sampaio (2010) é o das construções nas quais se verificam uma relação entre um possuidor humano e um objeto possuído; uma relação que, pelo menos a princípio, é do tipo tradicionalmente reconhecida como posse alienável, como (55) e (56).

(55) Rubinho quebrou o carro.

(56) A modelo quebrou o salto.

CAC-posse alienável pode ser interpretada em condições bastante específicas. Na verdade, o que temos na CAC-posse alienável é um tipo especial de conhecimento por parte do falante/ouvinte (por exemplo, saber que, em condições normais, o Rubinho não

quebraria seu próprio carro, e que a modelo não quebraria o próprio salto). Não havendo algum conhecimento prévio que imponha a interpretação ergativa, a interpretação desse tipo de estrutura tende a ser causativa, atribuindo ao sujeito o papel de agente e ao objeto o papel de paciente.

Sampaio (2010) argumenta que, em alguns casos, o que se percebe é que o sujeito da CAC-posse alienável representa metonimicamente o artefato diretamente atingido pelo dano, como (56), em que temos “modelo” pelo “sapato da modelo”, ou seja, a metonímia POSSUIDOR PELO POSSUÍDO. No caso (55), não há metonímia. A interpretação ergativa decorre da relação quase de posse inalienável que o falante/ouvinte estabelece entre Rubinho e seu carro de corrida; é como se o carro fosse conceptualizado como uma parte de Rubinho. Isso porque, todos sabemos, o carro é o instrumento de trabalho de Rubinho e o piloto é extremamente prejudicado quando seu carro sofre qualquer dano. Sampaio (2010) pontua ainda que a leitura ergativa não se mantém, ou pelo menos a interpretação da sentença se torna ambígua, quando o sujeito é Rubinho, mas o segundo SN não é um carro como em “Rubinho quebrou o barco”. Para a autora a CAC-posse alienável apresenta restrições bem específicas, que se refletem na sua frequência de uso. Como sua correta interpretação é mais dependente de informações contextuais do que as outras duas versões da CAC, seu uso acaba sendo mais limitado.

A CAC-posse alienável foi modelada no *Constructicon* da mesma maneira que as demais construções, de acordo com suas características semânticas e sintáticas. Como toda a família CAC, essa construção em questão, apresenta a mesma constituição sintática e a principal característica da CAC-posse alienável decorre de fatores pragmáticos, a modelagem dessa construção no modelo de língua utilizado por nós não é suficiente para abordar de forma satisfatória os construtos licenciados por essa construção. Para tratar desse fenômeno seria necessário um modelo de conhecimento de mundo, o que não é objeto desse trabalho. Dessa forma, nos experimentos que serão apresentados nos próximos capítulos, a CAC-posse alienável não estará incluída. Na Figura 70 temos a CAC-posse alienável no *Constructicon*.

Figura 70 – Construção Argumento Cindido-posses alienável

Argumento_cindido_posse_alienável [Argumento_cindido_posse_alienável] [131179]	
Definição	Tipo de construção de Argumento_cindido [SN[VISN]]. Essa construção exibe um argumento externo Sujeito e um Predicado. O argumento interno é um objeto direto. A construção apresenta uma relação entre um possuidor humano e um objeto possuído.
Exemplo(s)	
Elementos da Construção	<p>Predicado [Predicado] O Predicado é composto de um SV e um SN objeto direto.</p> <p>Sujeito [Sujeito] O sujeito é um SN.</p>
Relações	<p>Evoca Posse</p> <p>Herda de Argumento_cindido</p>

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

O último subtipo de construção da família CAC é a CAC-entidade/atributo, a construção envolve uma entidade e um de seus atributos, em predicados que expressam mudança em uma escala: aumentar/diminuir; melhorar/piorar, como em (57) e (58).

(57) A gasolina aumentou o preço.

(58) O cartão diminuiu o limite.

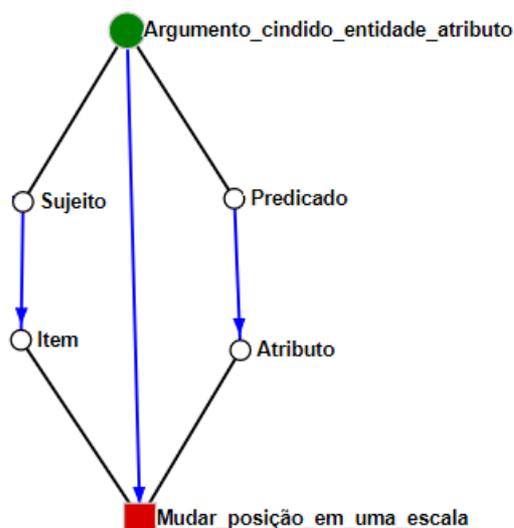
Nessa versão da CAC não há, como nos outros subtipos anteriormente descritos, uma restrição que remete a um dano sofrido por uma entidade. Na CAC-entidade/atributo, é evocado o frame de mudança escalar; ou seja, seus enunciados expressam uma mudança no posicionamento de um dos atributos de uma dada entidade em uma escala virtual.

Samapiao (2010) argumenta que, sintaticamente, temos um verbo inacusativo usado transitivamente. Semanticamente, as ocorrências apresentam a mudança de um sujeito-paciente. Além disso, a exigência de que exista uma relação de PARTE-TODO entre sujeito e objeto também é atendida, ainda que de uma forma especial. Nesse caso, os atributos de uma entidade são conceptualizados como partes dessa entidade ainda que metaforicamente.

Além do *frame* Ser_afetado via herança, temos ainda o *frame* Mudança_de_posição_em_uma_escala. Nesse *frame*, o EF Item é mapeado para o sujeito, a gasolina ou o cartão, o EF Atributo é mapeado para o objeto, o preço ou limite, e o EF Faixa_de_valor é mapeado para o verbo da construção, indicando a posição final da escala em que o Item termina.

Na Figura 71 temos a CAC-entidade/atributo com seus respectivos frames por ela evocados. Além dos *frames* Parte_Todo e Mudar_posição_em_uma_escala o *frame* Ser_afetado também é evocado por essa construção assim como nas outras, por isso, cada CE da construção está ligado a um EF de um *frame* específico.

Figura 71 – Frame evocado pela CAC-entidade/atributo



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Diante disso, Sampaio (2010) afirma que a Construção de Argumento Cindido é uma construção pragmaticamente motivada para evidenciar a função discursiva de tópico do SN-TODO. A CAC ocorre quando a entidade afetada é instanciada por um SN complexo no qual se observa uma relação do tipo PARTE/TODO. A alternativa que a língua oferece de imediato para a construção de SNs desse tipo é a expressão do TODO na forma de um adjunto adnominal: o pé do motorista; a tela do celular, o preço da gasolina.

As representações das CACs nos permitem perceber a semelhança com a representação da construção Transitiva, a diferença reside justamente na semântica dessas construções, enquanto na CAC os frames Ser_afetado e Parte_Todo, Partes_do_corpo e Mudar_de_posição_em_uma_escala estão unificados aos ECs Sujeito e Predicado, através dos EFs, a construção Transitiva indica que há um paciente que sofre a ação transitiva de um agente.

Quando comparamos as construções Ergativa, Argumento Cindido e Transitiva Direta Ativa, percebemos a necessidade de um sistema de representação que seja eficiente

para tratar de aspectos que vão além da estrutura formal. Se, por um lado, a CAC apresenta a mesma configuração sintática da Transitiva e por outro lado apresenta uma contraparte semântica semelhante a da Ergativa, cada construção possui aspectos específicos que devem ser considerados para que se produza uma representação eficaz dos fenômenos linguísticos.

Almeida (2016) pontua que o interesse em propor um tratamento linguístico-computacional para as três construções é devido à semelhança sintática ou semântica existente entre elas. A modelagem das construções conseguiu mapear que, de um lado, a CAC e a Ergativa são semanticamente sinônimas, mas com uma diferença entre elas, pois, enquanto a Ergativa apresenta uma estrutura monoargumental, em que uma entidade é afetada por um evento, a CAC apresenta uma estrutura de dois argumentos com a mesma informação semântica da Ergativa. Por outro lado, foi possível demonstrar como a CAC é semelhante à Transitiva Direta Ativa sintaticamente, devido a uma motivação que se manifesta pelo fato de ambas as construções herdarem de um tipo específico da construção Sujeito-Predicado, que especifica a transitividade do núcleo do Predicado como direta. (ALMEIDA, 2016, p. 90)

Dessa forma, percebemos a necessidade de um modelo que se baseie em pareamentos de forma e sentido em que se produza uma representação eficiente dos fenômenos linguísticos, de forma cognitivamente plausível. Para que no estágio posterior, a fase implementacional, tenhamos resultados mais satisfatórios.

5.10 CONSTRUÇÃO RESULTATIVA

Estudos recentes dentro do modelo teórico da Gramática das Construções e de outras abordagens têm apresentado reflexões acerca das idiosincrasias envolvendo a Construção Resultativa. Foi escolhida essa construção com o objetivo principal de proporcionar um tratamento construcionista computacional para esse fenômeno, diferenciando do tratamento recebido por estudos linguísticos anteriores.

A principal característica de uma Construção Resultativa em inglês é a existência de um Sintagma Adjetival que representa algum resultado de uma ação anterior. Segundo Boas (2003), a Construção Resultativa descreve o estado de um argumento que resulta de uma ação denotada pelo verbo.

Goldberg e Jackendoff (2004) propõem tratar as Resultativas do inglês como uma família de construções e não como um fenômeno específico.

A literatura no geral tem tratado as resultativas como um fenômeno unificado (o resultado). Nós pensamos que isso é um erro. Em nossos próprios trabalhos, nós tratamos as resultativas como uma família de construções (nós devemos chamá-las de subconstruções do resultado), compartilhando propriedades importantes, mas diferenciando em certas especificidades, incluindo seu grau de produtividade. (GOLDBERG & JACKENDOFF, 2004, p. 535)³

Ainda de acordo com Goldberg e Jackendoff (2004), uma sentença resultativa é caracterizada por um Sintagma Adjetivo ou Preposicionado que ocupa a posição de argumento verbal, assim o resultado aparece como um argumento sintagmático e não como um adjunto, como em (59) e (60).

(59) *Fred watered the plants flat.*

(60) *I cooked the carrots soft.*

Em (59), entende-se que, depois de aguadas, as plantas ficaram achatadas e, em (60), que as cenouras ficaram macias ao serem cozidas.

Ao observarmos esses exemplos, notamos que essas construções designam estados resultativos que estão contidos na ação descrita pelo verbo principal, ou seja, temos a descrição do estado de um argumento que deve ser entendido como um estado resultante causado pelo verbo.

Na abordagem construcional, Goldberg e Jackendoff (2004) defendem que o significado de uma construção resultativa é mais que a junção de dois eventos, uma vez que temos dois subeventos separados, um subevento verbal e um subevento construcional. Ao apresentar uma família de construções com essas características, Goldberg e Jackendoff (2004) abandonam a ideia rígida de que apenas o verbo sozinho determina a estrutura do sintagma verbal, por isso, descrevem as dimensões de variação nessas construções. A diversidade das construções resultativas em inglês, segundo Goldberg e Jackendoff (2004), abrange de construções transitivas e intransitivas a construções causativas e não causativas entre outras.

³ The literature on the whole has treated resultatives as a unified phenomenon ('the resultative'). We think this is a mistake. In our own work, we have treated resultatives as forming a sort of 'family' of constructions (we might call them subconstructions of the resultative), sharing important properties but differing in certain specifics, including their degree of productivity.

As Resultativas Transitivas são caracterizadas pelo Sintagma Resultativo ocupando a posição posterior ao objeto direto, já nas Resultativas Intransitivas, o Sintagma Resultativo ocupa a posição posterior ao verbo.

(61) *Herman hammered the metal flat.*

(62) *The pond froze solid.*

Os exemplos (61) e (62) mostram, respectivamente, as construções resultativas transitivas e intransitivas, em que o sintagma resultativo corresponde ao sintagma adjetival *flat* no caso transitivo e ao sintagma adjetival *solid* no caso intransitivo.

Outra variação das Resultativas do inglês, segundo os autores, ocorre quando temos a presença de um objeto reflexivo que não pode ser alternado com outro sintagma nominal, esse tipo é denominado de Resultativa Reflexiva Falsa, como podemos observar em (63), em que não pode haver outro pronome no lugar de *ourselves*.

(63) *We yelled ourselves hoarse.*

Há casos também em que o subevento construcional consiste no objeto direto tendo uma propriedade expressa pelo sintagma resultativo, as chamadas Resultativas de Propriedade como em (59), em que *flat* representa uma propriedade do subevento construcional.

De acordo com Goldberg e Jackendoff (2004), nos casos das Resultativas de Propriedade, o subevento verbal é o meio pelo qual o subevento construcional acontece. Logo, em (59), o evento verbal, que consiste em *Fred watered the plants*, permite que o subevento construcional aconteça, *the plants flat*, mostrando que a informação desse último é uma contribuição da construção e não do verbo. Em outras palavras, os autores afirmam que a estrutura argumental da semântica do subevento construcional determina a estrutura argumental sintática da sentença.

Em oposição às Resultativas de Propriedade, Goldberg e Jackendoff (2004) propõe as Resultativas de Trajetória, em que o subevento construcional consiste no sujeito fazendo uma trajetória expressa pelo sintagma resultativo, como em (64), em que o sintagma resultativo consiste em um sintagma preposicionado *down the hill* que expressa a trajetória realizada pelo sujeito *the ball*.

(64) *The ball rolled down the hill.*

Outra oposição apresentada pelos autores diz respeito às Resultativas Causativas e Não Causativas. Nas Resultativas Causativas, o subevento construcional consiste no sujeito causando no objeto direto a característica expressa pelo sintagma adjetival ou preposicionado como em (59). Já nas Resultativas Não Causativas, o subevento construcional é simplesmente uma mudança de estado ou mudança de posição do sujeito, como em (64).

As construções citadas apontam para a existência de uma variedade de Construções Resultativas em inglês, com propriedades semânticas ricas e de sintaxe diferenciada. Dessa forma, entendemos que uma análise léxico-sintática, puramente, é insuficiente para as Construções Resultativas, uma vez que, como Goldberg e Jackendoff (2004) afirmam, essas construções apresentam uma variedade de sintaxe atrelada a um significado especial. Abordagens meramente formais não darão conta das possibilidades de se expressar resultatividade em uma língua.

Além do trabalho de Goldberg e Jackendoff (2004), outros trabalhos significativos sobre resultativas na literatura serão expostos ao longo desse texto, mas, no momento, apresentaremos como essas construções têm sido tratadas no Português Brasileiro.

Existe um debate duradouro e produtivo acerca da existência de construções resultativas no PB. O trabalho de Foltran (1999) aparece como um dos pioneiros. Segundo a autora, as construções resultativas se aproximam do que normalmente se trata de predicado secundário resultativo. A autora parte do princípio de que as construções resultativas representam um verbo que causa uma mudança de estado no objeto que é denotado pelo sintagma adjetival. Nesse sentido, temos um caso de predicação secundária em que tanto o verbo como o adjetivo são responsáveis pela predicação.

Para a observação dos dados apresentados no trabalho de Foltran (1999) sobre as resultativas no PB, conclui-se que os dados são mais variados, atendendo o requisito de predicados secundários como observamos nos exemplos (65) a (68).

(65) Ela cortou o cabelo curto.

(66) Ela costurou a saia justa.

(67) Ele fabricou a cadeira torta.

(68) Ele cortou o pão em fatias.⁴

Como podemos observar nos exemplos, as construções resultativas apontadas por Foltran (1999) apresentam características semelhantes das resultativas do inglês apresentadas por Goldberg e Jackendoff (2004). Foltran (1999) argumenta que as construções resultativas de predicação secundária apresentam um sintagma resultativo que denota o estado atingido pelo referente do sintagma nominal do qual ele é predicado como um resultado da ação denotada pelo verbo. Por isso, em (65), o sintagma resultativo *curto* é composto por um sintagma adjetival que representa o resultado da ação verbal de cortar.

Já Barbosa (2018), argumenta que, no Português Brasileiro, não encontramos resultativas verdadeiras como acontecem em línguas como o inglês. Segundo ele, a análise realizada por Foltran (1999) apresenta o que ele denomina de pseudo-resultativas, uma vez que denotar estado resultante ou indicar resultado da ação não significa necessariamente que uma língua seja capaz de denotar algo semelhante a uma construção resultativa. Para ele, uma construção resultativa não é produto de mero acréscimo do predicado secundário sobre uma ação verbal. Construções resultativas são um fenômeno diferente dos recursos que uma língua como o PB possui para descrever essa estrutura de evento que aparecem após o verbo nos dados do PB e em línguas como o inglês que nos permite identificar resultativas genuínas.

Outros autores como Pustejovsky (1991) e Rappaport Hovav e Levin (1998) identificaram construções que foram nomeadas como resultativas verdadeiras e resultativas falsas para o inglês. No caso das resultativas verdadeiras, o sintagma resultativo descreve um evento independente, não lexicalmente especificado pelo verbo, enquanto, nas resultativas falsas, o sintagma final não o descreve, mas modifica o estado final especificado lexicalmente pelo verbo. Assim temos exemplos como (59) – reapresentado aqui como (69) para facilitar a leitura – que mostram casos de resultativas verdadeiras e (70), de resultativas falsas.

(69) *Fred watered the plants flat.*

(70) *John cut the bread into thick slices.*

⁴ Exemplos retirados de Foltran (1999)

Barbosa (2018) apresenta a seguinte definição sintática das construções resultativas genuínas: são estruturas mono-oracionais e transitivas formadas por um verbo que denota o modo da ação, um objeto direto afetado pela ação verbal (tema), e um predicado secundário que acrescenta à estrutura o estado resultante e a telicidade da ação.

Devido à falta de consenso sobre o verdadeiro status dos dados do PB perante a definição tradicional do que seja uma construção resultativa, neste trabalho, apresentamos o argumento segundo o qual definições baseadas em aspectos exclusivamente sintáticos são insuficientes para lidar com aspecto resultativo nas línguas em geral. Não nos interessa encontrar uma equivalência perfeita para o PB deste fenômeno visto em línguas como o inglês. O que propomos é demonstrar que, por mais que os dados do PB não se enquadrem estritamente nesse tipo de fenômeno, conforme descrito para o inglês, a semântica das resultativas encontra alguma maneira de ser expressa no PB.

Partindo do que propuseram Goldberg e Jackendoff (2004), que sugerem uma abordagem construcional em que uma interpretação está associada a uma estrutura sintática, uma vez que as línguas permitem os mesmos significados verbais apareçam em construções distintas, nossa proposta foi a de observar a relação de reciprocidade existente entre a semântica da construção e a semântica do verbo. Assim, a estrutura sintática estaria relacionada ao significado da construção, que corresponde ao subevento construcional e questões semânticas seriam consideradas para o melhor entendimento das construções.

Com vistas a propor a modelagem das construções resultativas, pretendemos apresentar uma análise diferenciada desse tipo de construção. Diferentemente da classificação realizada por Foltran (1999), que não leva em conta as diversas nuances de sentido que pode adquirir um verbo dependendo das relações que se estabelecem no interior das construções, propomos uma análise que leve em conta a semântica da construção e a semântica do verbo, mapeando assim como o significado construcional contribui para a existência de resultativas no Português Brasileiro.

Para esse fim, temos:

- (i) em português, há construções que podem ser consideradas resultativas, ainda que não haja exemplos das resultativas consideradas prototípicas;
- (ii) As construções mais produtivas, quando se trata de expressar resultatividade em português, são as que apresentam a forma sintática [SN

V SN SAdj], tendo como verbos relacionados àqueles de ação em que há uma relação específica com o objeto.

Com esse aporte teórico-metodológico, empreende-se, então, a apresentação de uma proposta de ir além descrição sintática até então feita nos estudos sobre resultativas para o PB propomos classificar as construções em estudo com base na semelhança familiar na sintaxe e semântica com a Resultativa transitiva de propriedade causativa (GOLDBERG; JACKENDOFF, 2004) e apresentar a modelagem dessas construções no *Constructicon*.

Diante das considerações apresentadas, os construtos que apresentam característica resultativa além da estrutura sintática [SN [SV SN [SAdj]]] também evocam o *frame* Causar_mudança. A Construção Resultativa no PB foi modelada conforme essas propriedades, uma vez que nesse *frame* temos a existência de um agente que provoca uma alteração em uma entidade em sua categoria. A Figura 72 mostra a modelagem dessa construção.

Figura 72 – Construção Resultativa

Resultativa [Resultativa]	
Definição	Tipo de construção [SN [V SN[SAdj]]]. Essa construção exibe um argumento externo Sujeito e um Predicado complexo em que uma ação verbal afeta um Tema e o resultado dessa ação aparece no SAdj.
Exemplo(s)	
Elementos da Construção	<p>Predicado [Predicado] O Predicado por um Sintagma verbal complexo em que temos um SN e um SAdj.</p> <p>Sujeito [Sujeito] O Sujeito é um SN agentivo.</p>
Relações	<p>Evoca Causar_mudança</p> <p>Herda de Sujeito_predicado</p>

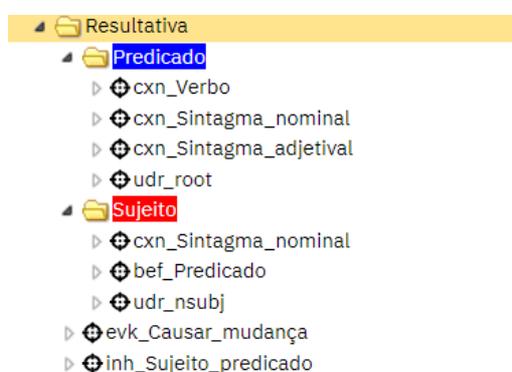
Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Ao considerarmos a existência da Construção Resultativa no PB, o *frame* Causar_mudança, evocado por ela, apresenta como EFs nucleares, o Agente, a Entidade, a Categoria Inicial e a Categoria Final, sendo cada EF unificado a um CE da construção. O Sujeito, que é composto por um SN, é unificado ao Agente. Como temos aqui um Predicado complexo, que é composto de um verbo, um SN e um SAdj, a Entidade e a Categoria Inicial são unificadas ao SN, e o resultado representado pela Categoria Final é unificado ao Sintagma Adjetival.

Da mesma forma que ocorre com a Construção Transitiva Indireta, a Construção Resultativa, também é construída a partir da Proeminência (LANGACKER, 1991): neste caso o elemento mais proeminente é a Entidade em relação a Categoria Inicial. Em “Maria cortou o cabelo curto” o SN “o cabelo” representa tanto a Entidade quanto a Categoria Inicial, já que podemos inferir que o “cabelo” estava maior antes de passar pelo processo de corte. Dessa forma, o *frame* Causar_mudança apresenta o perfilamento da Entidade em relação a Categoria Inicial, uma vez que este último pode estar incorporado a própria Entidade.

A Figura 73 mostra como essas restrições foram aplicadas aos ECs da construção com intuito de proporcionar um tratamento adequado para a interpretação dessa construção. Já a Figura 74 mostra como o *frame* evocado foi mapeado para os ECs da Construção Resultativa. Podemos observar que a nossa proposta de considerar, principalmente, o aspecto semântico da Construção Resultativa afim de exemplificar como essas estruturas se manifestam no PB foi possível devido a identificação que a fenômeno considerado como resultativo em outras línguas também é caracterizado pela semântica de Causar_mudança.

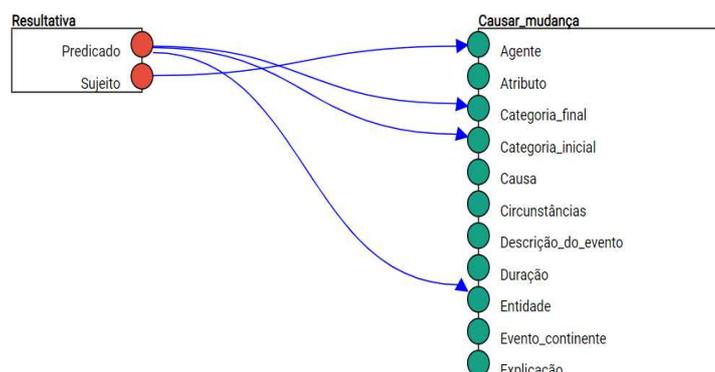
Figura 73 – Restrições aplicadas aos ECs da Construção Resultativa



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 74 – Relação Evoca CAC-parte do corpo

Relation type: Evoca



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.11 AS CONSTRUÇÕES PREDICATIVAS

Nesta seção apresentamos mais um conjunto de construções de estrutura argumental do PB modeladas até então no Constructicon da FrameNet Brasil: as construções Predicativas, chamadas por Perini (2010) de Estativas, por conta do fato de o *slot* V nessas construções ser preenchido por verbos de estado, tais como *ser*, *estar*, *continuar*, *ficar* e *parecer*. Segundo este autor (PERINI, 2010, p. 101-102), há dois subtipos de construções Estativas: a Estativa propriamente dita, em que uma Qualidade é atribuída a um Qualificando, e a Estativa de Lugar, em que um Lugar é atribuído a um Localizando. As sentenças (71) e (72), extraídas de Perini (2010), exemplificam essas construções.

(71) A Letícia é muito inteligente.

(72) A Letícia é um gênio.

Analisando-se (71) e (72), nota-se que o *slot* do predicativo pode ser preenchido tanto por um núcleo adjetival, quanto por um núcleo nominal.

A análise de Perini (2010) é adaptada à modelagem dessas construções no *Constructicon* da FrameNet Brasil, as quais, porém, são nomeadas como construção Predicativa Nominal e construção Predicativa Locativa. Tal nomeação decorre da constatação de que o esquema [SN[Vcop[SN|SAdj]]], que caracteriza a construção Estativa de Perini (2010) se presta a relacionar uma entidade – o Sujeito – tanto a um

estado, quanto a um atributo. Assim, há dois subtipos de construção Predicativa Nominal: a Atributiva, que licencia sentenças como (71), e a Estativa, que licencia casos como (73).

(73) O peixe está doente.

Castilho (2010) apresenta uma nomenclatura diferente de Perini (2010), já que todos os exemplos acima são nomeados por ele de Construções Estativas, segundo o autor, nessas construções os *slots* dos verbos são bem restritos e devem ser preenchidos por *ser, estar, permanecer, ficar, tornar-se, parecer* e poucos mais. A estrutura sintagmática apontada por Castilho (2010) é a mesma apresentada por Perini (2010), e, para o primeiro, a predicação nessas construções se concentra no sintagma nominal ou no sintagma adjetival que seguem ao verbo.

A existência dos verbos chamados copulares nas Construções Predicativas contribui para esse caráter atributivo e estativo delas. Ao revisitar gramáticas tradicionais (CUNHA & CINTRA, 2013; ROCHA LIMA, 2007), constata-se, mais uma vez, a definição do verbo de ligação ou cópula como elo entre sujeito e predicativo, além da apresentação de uma lista com os verbos prototípicos (*ser, estar, permanecer, ficar, tornar-se*) e seus respectivos estados (*permanente, transitório etc.*). Para Raposo et al. (2013) esses verbos podem ser classificados semanticamente em dois grupos: aqueles que são usados para atribuir uma propriedade ao sujeito ou descrever seu estado, e aqueles que descrevem uma mudança de estado do sujeito. Assim, identifica os verbos que se encaixam em cada grupo. O autor também faz uma distinção entre predicados estáveis e predicados episódicos, que diz respeito às propriedades mais permanentes e às mais transitórias de um indivíduo na configuração da predicação nominal. Embora haja um consenso na literatura sobre o fato de verbos de cópula não contribuírem semanticamente para o conteúdo do predicado, a partir do momento em que tais verbos marcam determinadas orientações, há, sim, uma contribuição semântica para a predicação.

As Construções Predicativas Nominais enquadram-se, nas predicções de base nominal/adjetival, em que não é o verbo que veicula o conteúdo daquilo que se diz sobre o sujeito, mas sim o SN e o SAdj. Diante disso, nossa proposta de divisão implica na Construção Predicativa Nominal Atributiva que evoca o *frame* de Atributos, e a Predicativa Nominal Estativa evoca o *frame* de Estado.

A Predicativa Nominal Atributiva, que indica um atributo ligado a um sujeito, aceita todos os verbos de ligação listados por Perini (2010) e foi modelada conforme as

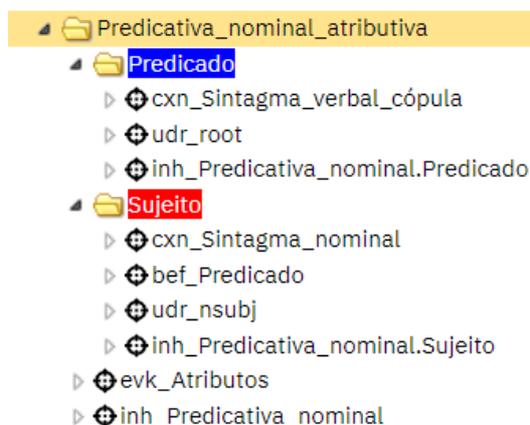
Figuras 75 e 76 que demonstram a construção juntamente com as restrições aplicadas aos ECs.

Figura 75 – Construção Predicativa Nominal Atributiva

Predicativa_nominal_atributiva [Attributive_nominal_predicate]	
Definição	Tipo de construção predicativa em que um SAdj ou SN codificam um atributo do Sujeito .
Exemplo(s)	
Elementos da Construção	<p>Predicado [Predicate] O Predicado é composto por um SAdj ou SN que codificam um atributo do Sujeito.</p> <p>Sujeito [Subject] O Sujeito é um SN.</p>
Relações	<p>Evoca Atributos</p> <p>Herda de Predicativa_nominal</p>

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 76 – Restrições aplicadas aos ECs da Construção Predicativa Nominal Atributiva



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Além disso, restrições aplicadas sobre os verbos de ligação, ou cópulas, que podem figurar em cada subtipo, outro fator importante para a modelagem foi a criação de uma construção Predicativa Nominal mais genérica e abstrata que não licencia nenhum construto específico, mas fornece estrutura para as demais predicativas.

A Construção Predicativa Nominal Estativa não aceita o verbo *ser*, uma vez que a noção de permanência advinda desse verbo não condiz com a noção de estado. Temos então, a relação de herança da Predicativa Nominal Estativa, com a Predicativa Nominal genérica, juntamente com a relação Evoca em que o *frame* de Estado, descreve a relação

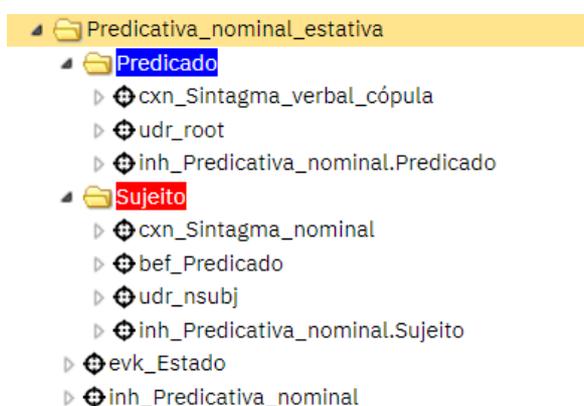
que o Predicado tem com o Sujeito, uma vez que, um atribuí estado ao outro. Essas informações podem ser vistas nas Figuras 77 e 78, em que temos a Construção Predicativa Estativa e as restrições que foram aplicadas aos ECs da construção.

Figura 77 – Construção Predicativa Nominal Estativa

Predicativa_nominal_estativa [Stative_nominal_predicate]	
Definição	
Tipo de construção [SN[Vcop[SAdj]SN]] em que um adjetivo ou nome predicativos atribuem um estado a um Sujeito .	
Exemplo(s)	
Elementos da Construção	
Predicado [Predicate]	O Predicado é composto por um SAdj ou SN predicativos.
Sujeito [Subject]	O Sujeito é um SN.
Relações	
Evoca	Estado
Herda de Predicativa_nominal	

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 78 – Restrições aplicadas aos ECs da Construção Predicativa Nominal Estativa



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Além das Predicativas Nominais, há ainda a Construção Predicativa Locativa que apresenta o esquema sintático [SN [V_{cop} [SAdv|SP]]] e indica a localização de uma entidade, licenciando construtos como em (74) e (75).

(74) A toalha está aqui.

(75) Poços de Caldas é no Sul de Minas.

Diferentemente das Predicativas Nominais, a Predicativa Locativa não indica um atributo ou um estado do sujeito. Pelo fato de os verbos serem semelhantes, o caráter locativo se dá por uma contribuição construcional: uma vez que esses verbos podem aparecer em instâncias sintáticas distintas, a diferença de significado resulta de semânticas construcionais distintas. Como afirma Goldberg (1995), o sentido global da construção resulta da associação entre o significado da construção e o significado dos itens lexicais que a compõem. Por isso, enquanto nas Predicativas Nominais o predicativo do sujeito deve ser constituído de núcleos nominais ou Sintagmas adjetivais, já em (74) e (75), esse mesmo *slot* pode ser preenchido tanto por núcleos adverbiais, quanto por Sintagmas Preposicionais. A Construção Predicativa Locativa foi modelada conforme a Figura 79, em que temos sua descrição de acordo com sua constituição, juntamente com suas relações de Herança e Evoca, indicando que essa construção é herdeira da Sujeito_Predicado e evoca o *frame* Ser_localizado.

Figura 79 – Construção Predicativa Locativa

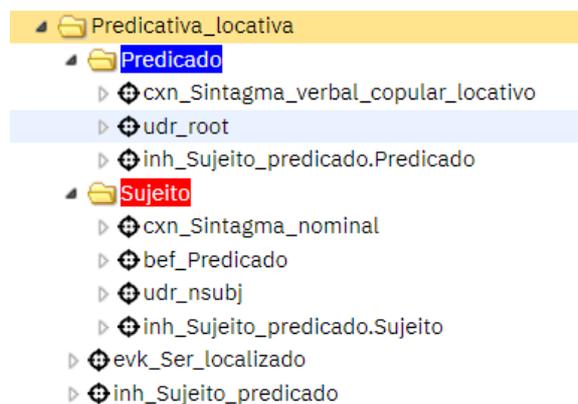
Predicativa_locativa [Locative_predicate]	
Definição	Tipo de construção em que o núcleo é um advérbio ou nome Predicativo que atribui uma localização a um Sujeito .
Exemplo(s)	
Elementos da Construção	<p>Predicado [Predicate] O predicado é composto por uma cópula e um SAdv ou SP.</p> <p>Sujeito [Subject] O Sujeito é um SN.</p>
Relações	<p>Evoca Ser_localizado</p> <p>Herda de Sujeito_predicado</p>

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

A informação sobre a constituição do EC Predicado é modelada aplicando-se restrições tanto ao EC em si, quanto à construção que o licencia. Assim, utilizando-se a restrição CE>Construction, modela-se que o EC Predicado é licenciado pela construção de Sintagma_verbal_copular_locativo, como demonstra a Figura 80, a qual, por sua vez, é composta pelos ECs Cópula e Locativo. Em seguida, aplicam-se ao EC Cópula as restrições CE>Construction, para indicar que ela é licenciada pela construção de Verbo; CE_before, para indicar que a Cópula vem antes do Predicado, e CE>Lexeme, para

indicar que apenas os verbos *ser*, *estar*, *ficar* e *continuar* podem ser usados nesse *slot*, excluindo-se o verbo *parecer*, conforme aponta Perini (2010).

Figura 80 – Restrições aplicadas aos ECs da Predicativa Locativa



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.12 CONSTRUÇÕES VERBOS-NOMINAIS

Quando falamos de Construções Verbos-Nominais há uma gama de estudos acerca dessas construções. Alguns desses estudos tratam as construções com verbos suporte como Construções Verbos-Nominais que seriam o resultado da associação de um verbo semanticamente leve com um sintagma nominal. No entanto, nossa proposta, neste trabalho, não é lidar com essas estruturas com verbo suporte. As Construções Verbos-Nominais apresentadas aqui dizem respeito às construções que, segundo Castilho (2010), são organizadas em termos do verbo e seus complementos, formando uma sentença simples, e sintagmas nominais não selecionados pelo verbo formando uma *minissentença*.

Castilho (2010) apresenta um estudo sobre o que ele denomina de minissentença. De acordo com o autor, as minissentenças são tratadas na Gramática Tradicional como Predicados Verbo-Nominais, uma vez que o verbo ultrapassa as relações transitivas e escolhe simultaneamente um objeto e um complemento de qualidade. Assim, temos um predicado que é verbal, porque escolhe um objeto direto, e é nominal, porque encaixa um adjetivo predicativo no conjunto.

Seguindo os pressupostos de Castilho (2010) as Construções Verbo-Nominais presentes nesse trabalho – denominadas por ele de Minissentenças Adjetivais – licenciam construtos como aqueles em (76) e (77).

(76) Maria considerou o presidente um imbecil.

(77) João apareceu bêbado.

Como podemos perceber, essa construção apresenta uma predicação adjetival que é expressa pelo predicado verbo-nominal. Segundo Cunha e Cintra (2013), o predicado verbo-nominal em português possui dois núcleos significativos, um verbo e um predicativo. Já Castilho (2010) afirma que nas minissentenças adjetivais temos uma única estrutura desempenhando duas funções, o que evidencia o multifuncionalismo das estruturas sintáticas. No exemplo (78), retirado de Castilho (2010), o autor afirma que *doente* é uma minissentença adjetival que se encontra separada do sintagma verbal *encontraram o povo*, havendo, assim, uma fronteira sintática entre a sentença simples e a minissentença adjetival.

(78) Os pesquisadores encontraram o povo doente.

Diante dos apontamentos expostos por Castilho, as Construções Verbo-Nominais serão tratadas sobre dois aspectos: o subjetivo, que engloba construtos como (77), e o objetivo, com construtos como (76).

A Construção Verbo-Nominal Objetiva, nomenclatura dada por nós, para as minissentenças adjetivais apresentadas por Castilho (2010), também foi estudada por Foltran (1999) sobre o mesmo viés de predicação secundária da Gramática Tradicional. Segundo a autora, esse tipo de estrutura é denominada de mini oração em que o segundo argumento do verbo, o objeto direto, é expresso por uma oração, como em (76), e em (79) em que “o presidente um imbecil” e “Maria bonita” constitui, segunda a autora, uma mini oração de predicação não verbal, sendo o adjetivo o predicador do sintagma nominal.

(79) Eu considero Maria bonita.

A Construção Verbo-Nominal Objetiva é um tipo de construção de estrutura argumental que apresenta a seguinte configuração [SN[SV [SN[SAdj]]]], havendo, assim, a existência de outro sintagma nominal entre o verbo e sintagma adjetival.

As Construções Verbo-Nominais foram tratadas na literatura, basicamente sob o aspecto da dupla predicação como vimos na abordagem de Foltran (1999). A Gramática Tradicional, igualmente, trata essas estruturas como predicados verbos-nominais. Segundo Cunha e Cintra (2013), o predicado verbo-nominal em português possui dois núcleos significativos, um verbo e um predicativo, afirmando que verbos plenos também podem construir predicativos do sujeito.

Como mencionado anteriormente, Castilho (2010) apresenta uma outra abordagem: segundo o autor, o rótulo de minissentença que corresponde, neste trabalho, às Construções Verbo-Nominais, possui uma associação com o inglês - *small clause* – e está baseado na convicção de que esta relação semântica reflete-se uniformemente na estrutura de constituintes, no sentido de que a relação Sujeito/Predicado é sempre codificada sintaticamente em termos de um par de constituintes irmãos, tal como S - sujeito + predicado.

A minissentença adjetival, neste caso a Verbo-Nominal Objetiva, sendo codificada em termos sintáticos através do Sujeito/Predicado, possui características diferentes da Verbo-Nominal Subjetiva. O adjetivo que a constitui, mesmo não sendo considerado como um constituinte do sintagma nominal e, sim, como uma unidade autônoma, deve ser entendido como uma contribuição da semântica construcional e não lexical, uma vez que a estrutura composicional contribui para o significado das sentenças, mas não é responsável por todo significado.

A Figura 81 exemplifica como a Construção Verbo-Nominal Objetiva foi modelada em termos de suas características sintáticas e sua relação com outras. Da mesma maneira que ocorreu nas Construções Predicativas, aqui também se fez necessária a criação de uma Construção Verbo-Nominal abstrata e mais genérica que fornece estrutura para a suas construções herdeiras.

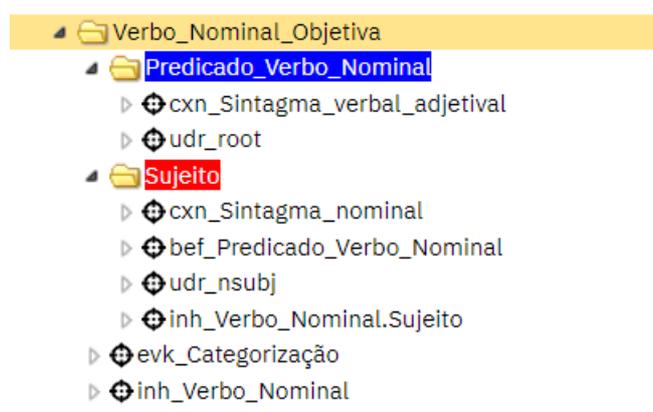
Figura 81 – Construção Verbo-Nominal Objetiva

Verbo_Nominal_Objativa [Verbo_Nominal_Objativa]	
Definição	
Tipo de construção [SN [V SN [SAdj]]] em que o núcleo é um Predicado verbo-nominal.	
Exemplo(s)	
Elementos da Construção	
Predicado_Verbo_Nominal [Predicado]	O predicado é composto por um Sintagma Verbo-Nominal.
Sujeito [Sujeito]	O Sujeito é um SN.
Relações	
Evoca	Categorização
Herda de Verbo_Nominal	

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

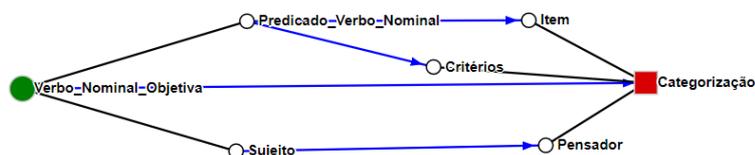
Já as Figuras 82 e 83 apresentam a construção Verbo Nominal Objetiva com as respectivas restrições modeladas. A partir das restrições vemos como a constituição da construção foi modelada, uma vez que temos uma construção composta por um Sujeito que é um sintagma nominal e um Predicado Verbo Nominal que consiste em um sintagma nominal e um sintagma adjetival. A contraparte semântica da construção foi modelada através do *frame* Categorização, que indica que um Item se encaixa em determinados Critérios conforme a construção de um Pensador. Nesse sentido, temos que o Pensador corresponde ao SN-sujeito e o Predicado Verbo Nominal corresponde aos EFs Item e Critérios no frame evocado pela construção.

Figura 82 – Restrições aplicadas aos ECs da Verbo_Nominal_Objativa



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Figura 83 – Frame evocado pela Construção Verbo_Nominal_Objativa



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Por fim, temos a Construção Verbo-Nominal Subjetiva, denominada por nós como Construção de Mudança de Estado⁵. Conforme já pontuado, a Construção de Mudança de Estado, que licencia (77), é um tipo de construção de estrutura argumental caracterizada por [SN[V[SAdj]]].

Com uma abordagem gerativista, Foltran (1999) apresenta essas construções como predicacões secundárias. Ela faz uma divisão entre predicacão secundária e mini-oração, sendo que a predicacão secundária pode ser orientada para o objeto, para o sujeito, ou ser caracterizada como resultativa. Segundo Foltran, sentenças como (80) são exemplos de predicados secundários orientados para o sujeito.

(80) Pedro chegou cansado.

O que caracteriza, de acordo com Foltran (1999), a predicacão secundária é o recebimento de um papel temático do sujeito ou objeto por um outro núcleo lexical, assim temos duas unidades predificadoras. Em (80), *Pedro* é duplamente tematizado: pelo verbo chegar e pelo adjetivo cansado.

Como podemos observar, a análise realizada por Foltran apresenta um caráter mais composicional e lexicalista, uma vez que essa construção é apresentada apenas como resultado de uma dupla tematização de dois predicados.

As descrições realizadas por Castilho e Foltran demonstram casos em que a construção de mudança de estado apresenta a seguinte configuração [SN [V [SAdj]]],

⁵ O estudo da Construção de Mudança de Estado foi submetido e publicado na Revista Linguística. (DINIZ DA COSTA ET AL, 2018) Representação computacional das construções de sujeito-predicado do português do Brasil. Linguística, Rio de Janeiro, v. 14, p. 149.
DOI: <http://dx.doi.org/10.31513/linguistica.2018.v14n1a15285>

por isso, nesse trabalho para a modelagem de construções, tratamos de construtos como (77).

Nosso objetivo aqui é, primeiramente, demonstrar como uma abordagem construcional pode ser mais eficaz para tratar de estruturas como essas, pois, como afirma Goldberg (1995), estruturas argumentais são instâncias de construções que existem independentemente de verbos – ou outros itens lexicais predicadores, como adjetivos – e evocam frames genéricos das ações humanas básicas, possuindo, assim, significado por si mesmas, o que nos permite diferenciar uma construção de outra. Em segundo lugar, pretendemos ainda demonstrar a possibilidade de modelar tal construção computacionalmente utilizando a infraestrutura disponibilizada pela FrameNet Brasil. Considerem-se os exemplos (81) a (84), extraídos do Corpus do Português – Web/dialetos.

- (81) Um belo dia, Jack **amanheceu morto**.
- (82) João **acabou condenado** apenas por homicídio culposo.
- (83) No último show, em Belgrado, na Sérvia, mês passado, Amy **apareceu bêbada**, esqueceu letras e saiu do palco antes da hora.
- (84) O Japão **acordou feliz** e muito entusiasmado por organizar a maior competição esportiva do mundo.

Apesar de os itens lexicais que ocupam os slots do verbo e do adjetivo na construção serem distintos, eles guardam algo em comum. Começando pelos verbos, assim como *amanhecer*, *acabar*, *aparecer* e *acordar* não preveem, em sua especificação de valência (val) um argumento interno, pedindo apenas um sujeito. Quanto aos adjetivos, assim como *doente*, *morto*, *condenado*, *bêbado* e *feliz* indicam estados. Nesse sentido, as restrições aplicadas à modelagem consideram que o sujeito da construção seja unificado à Entidade a quem o Estado se aplica, ou de que o frame evocado pelo verbo seja unificado a uma determinada Circunstância da mudança de estado. Dados os princípios e ferramentas de modelagem já descritos, a construção de Mudança de Estado pode ser definida no Constructicon da FrameNet Brasil conforme a Figura 84.

Figura 84 – Construção Mudança de Estado

Mudança_de_estado [state_change]	
Definição	
Tipo de construção [SN[V[SAdj]]]. Essa construção é composta por um SN Sujeito e um Sintagma Verbal Adjetival Predicado_verbo_nominal . A construção indica uma mudança de estado no Sujeito.	
Exemplo(s)	
Elementos da Construção	
Predicado_Verbo_Nominal [predicado verbo nominal]	Predicado composto obrigatoriamente por um Sintagma_verbal_adjetival.
Sujeito [state_change_sujeito]	O sujeito é um Sintagma_Nominal.
Relações	
Evoca	Sofrer_mudança
Herda de	Sujeito_predicado

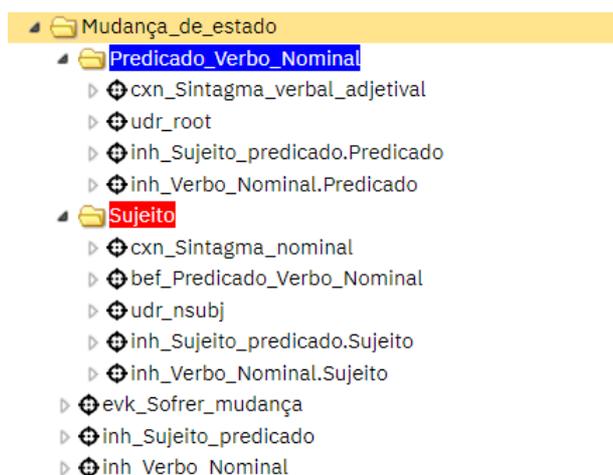
Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Como se pode notar pelas descrições, os Elementos da Construção (ECs) são licenciados pelas construções de Sintagma Nominal e Sintagma Verbal Adjetival, que correspondem ao Sujeito e ao Predicado Verbo-Nominal, respectivamente. Para garantir que essa informação conste da representação computacional dessa construção, a restrição CE>Cxn (cxn) é aplicada.

Ademais, fez-se necessário indicar que o *frame* evocado pelo Adjetivo deve evocar algum Estado. Assim, a restrição CE>Frame_family (fam) é aplicada ao núcleo Adjetivo do CE Predicado verbo nominal. Por fim, fez-se necessário restringir os verbos que ocorrem no slot V àqueles que são admitidos pela construção. Dois tipos de restrições podem ser utilizadas para tanto: a restrição CE>Frame, para os casos em que todos os verbos que evocam um dado frame podem ocupar o slot; ou a restrição CE>LU, quando se quiser definir verbos específicos, com sentidos específicos, a ocupar esta posição.

A aplicação de tal restrição demandaria um estudo na fase linguística que não está no escopo desse trabalho, assim, a título de demonstração, aplicamos a restrição CE>Frame (evk) ao V, para indicar que quaisquer verbos que evoquem o frame de Acordar – acordar, despertar – podem preencher essa posição. O resultado da aplicação das restrições é mostrado na Figura 85.

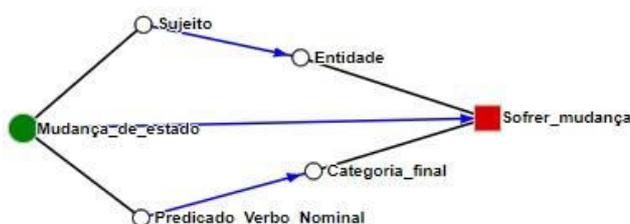
Figura 85 – Restrições aplicadas aos ECs da Construção Mudança de Estado



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Completando a modelagem, temos por um lado a relação de Herança, já que a construção de Mudança de Estado da construção mais genérica Verbo-Nominal sendo que, naquela, o CE Predicado é mais específico do que aquele da construção mãe, por outro a relação de Evocação representada na figura 86, em que seu aporte semântico é modelado para o *frame* de Sofrer_mudança através do mapeamento dos ECs Sujeito e Predicado verbo nominal aos EFs Entidade e Qualidade_final, respectivamente.

Figura 86 – Frame evocado pela Construção Mudança de Estado



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

5.13 CONSTRUÇÕES CLAUSAIS SEM SUJEITO

Esta seção, diferentemente das outras seções, não apresenta construções herdeiras da construção Sujeito_Predicado. O que propomos aqui é a descrição de um tipo de estrutura muito presente no PB: as chamadas orações sem sujeito. De acordo com a gramática tradicional, orações sem sujeito podem ser definidas como estruturas em que

interessa apenas o processo verbal sem atribuição de nenhum ser a esse processo, afirmando-se a impessoalidade do verbo e a inexistência do sujeito (CUNHA E CINTRA, 2013).

De acordo com Cunha Cintra (2013), os principais casos de inexistência do sujeito ocorrem em: verbos e expressões que denotam fenômenos da natureza, verbo *haver* na acepção de existir, verbos *haver*, *fazer* e *ir* quando indicam tempo decorrido e o verbo *ser* na indicação de tempo em geral como em (85) a (88).

- (85) Choveu.
- (86) Há três quadros na sala.
- (87) Faz muito frio em julho.
- (88) Era noite quando cheguei.

Rocha Lima (2011) também afirma que as orações sem sujeito se esgotam no processo verbal em si, sendo expressas através de verbos impessoais, ou seja, que exprimem algo que não pode ser atribuído a nenhuma pessoa gramatical.

Castilho (2010) apresenta uma perspectiva e, conseqüentemente, uma nomenclatura diferente das expostas por Cunha e Cintra (2013) e Rocha Lima (2011): o autor afirma que estruturas em que se indicam fenômenos da natureza são denominadas de sentenças não argumentais, constituídas por verbos não argumentais que são simultaneamente impessoais, não dispondo, assim, nem de sujeito nem de argumento interno, como em (85).

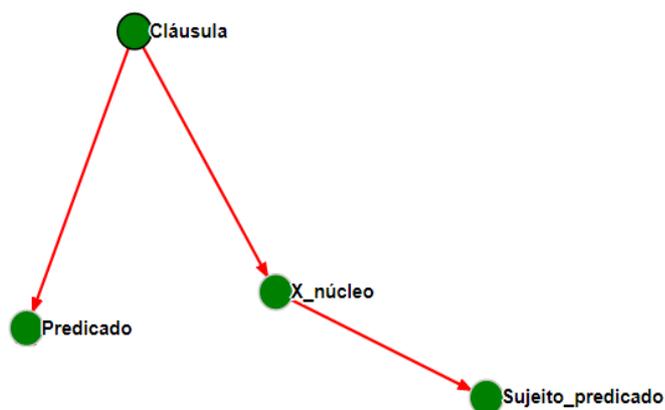
Os outros exemplos apontados pela Gramática Tradicional como orações sem sujeito são denominados por Castilho (2010) como sentenças apresentacionais ou existenciais. Segundo o autor, essas sentenças são monoargumentais e funcionam no discurso para a introdução de um tópico novo através da estrutura sintática [V SN]. Para Castilho (2010), sentenças como (86) a (88) são caracterizadas pela presença de um verbo impessoal, frontado que organiza a estrutura sintagmática (verbo + sintagma nominal), e esse verbo não atribui papel temático, com isso, a predicação não é de responsabilidade dele, mas das expressões locativas e temporais que coocorrem. Nessas estruturas, o sintagma nominal único é geralmente indefinido e vem habitualmente posposto ao verbo salvo quando ele é topicalizado. Castilho (2010) argumenta também que as sentenças monoargumentais podem ser consideradas implicitamente como locativas, pois, para serem devidamente interpretadas, a afirmação de que alguma coisa existe ou existiu

requer complementação com uma expressão locativa ou temporal. Em nossos exemplos o locativo ocorre em (86) *na sala*, a expressão de tempo em (87) *em julho*.

A partir das considerações apontadas para este trabalho, consideramos que as orações sem sujeito, apresentadas por nós como Construções Clausais Sem Sujeito, são estruturas informacionais que representam um fato propriamente dito sem referência a qualquer ser ou entidade – fenômenos naturais –, como também constatações referentes a algo ou alguém – função apresentativa.

A modelagem dessas construções no *Constructicon* definiu a existência de uma construção denominada Cláusula, de caráter abstrato, da qual herdam a Construção X_Núcleo e a Construção Predicado. A Figura 87 demonstra como a relação de Herança mapeia que a Construção Cláusula, do mesmo modo que fornece estrutura para a Construção Sujeito_Predicado, fornece para a Predicado.

Figura 87 – Construções Herdeiras de Cláusula



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Para tratarmos das instanciações das Construções Clausais sem Sujeito, entendemos que, por essas construções não apresentarem argumento externo, são herdeiras da Construção Predicado. Nesse trabalho, as Construções Clausais sem Sujeito se encontram divididas em: Construção Sem Sujeito Intransitiva, Construção Existencial Apresentacional e Construção de Tempo Decorrido. O primeiro tipo abarca construtos como (89) e (90).

(89) Anoiteceu.

(90) Ventou.

A Construção Sem Sujeito Intransitiva que licencia (89) e (90) apresenta uma estrutura não argumental composta apenas por um Verbo de caráter impessoal que denota um fenômeno da natureza. O CE dessa construção, o Predicado, apresenta como restrição um verbo que não seleciona nenhum argumento interno, tampouco argumento externo. A Construção Sem Sujeito Intransitiva foi modelada conforme a Figura 88 em que podemos observar a definição da construção com seu respectivo CE e sua relação de Herança com a Construção Predicado.

Figura 88 – Construção Sem sujeito Intransitiva

Sem_sujeito_intransitiva [Sem_sujeito]	
Definição	Construção que denota fenômenos da natureza. Composta por um Verbo_impessoal .
Exemplo(s)	
Elementos da Construção	Verbo_impessoal [verbo_impessoal] O núcleo verbal preenchido pelo verbo impessoal
Relações	<ul style="list-style-type: none"> Evoca Fenômenos_naturais Herda de Predicado

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

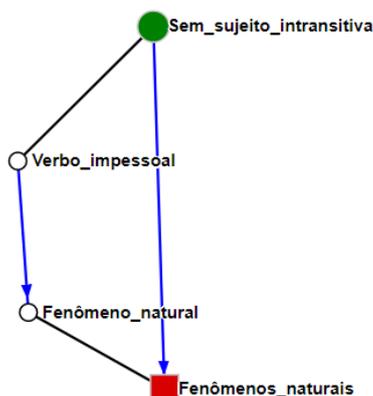
Algo que vale ressaltar quanto aos verbos que designam fenômenos da natureza é que eles podem ocorrer em construções que não são da tipologia sem sujeito. No exemplo (81), vemos o construto *Jack amanheceu morto*, em que o verbo em questão, apesar de ser considerado um verbo não argumental, está configurado na Construção Mudança Estado, reforçando a premissa que o significado é, em parte, independente dos componentes lexicais de uma estrutura. O exemplo (81) evidencia que a contribuição argumental é da construção e não do verbo, assim a diferença entre (81) e (91) resulta de semânticas construcionais distintas.

(91) Amanheceu.

Dessa forma, para representar o significado da Construção Sem Sujeito Intransitiva utilizamos o *frame* Fenômenos_naturais indicando um acontecimento natural que ocorre sem interferência humana. Por se tratar de uma construção com apenas um EC que

corresponde ao verbo impessoal responsável por indicar o fenômeno natural, apenas o EF nuclear do frame em questão se encontra unificado ao EC como podemos observar na Figura 89.

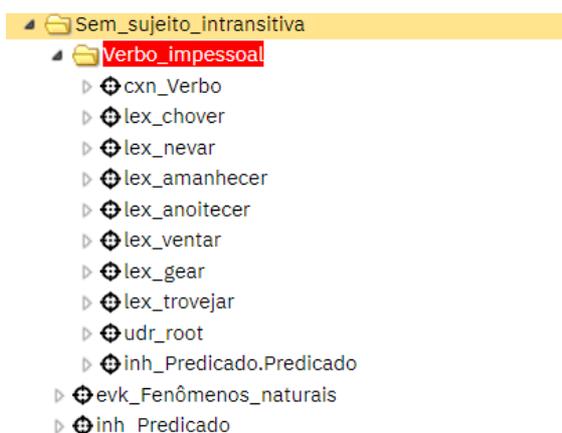
Figura 89 – Frame evocado pela Construção Sem_Sujeito_Intransitiva



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Para uma representação mais eficiente da Construção Sem_Sujeito_Intransitiva foram modeladas restrições que dizem respeito a algumas características da construção. O exemplo presente na Figura 90 mostra como restrição CE>lexeme foi utilizada para especificar os principais verbos que indicam fenômenos naturais. O intuito dessa restrição serve para indicar que apenas esses verbos podem ocupar esse *slot* na construção.

Figura 90 – Restrições aplicadas aos ECs da Construção Sem_Sujeito_Intransitiva



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

A Construção Existencial Apresentacional constitui outro tipo de construções clausais sem sujeito. Essa construção apresenta a configuração sintática [V SN] em que o verbo também é do tipo impessoal como em (92) e (93).

(92) Há um carro ali.

(93) Tem um gato na cozinha.

Os exemplos (92) e (93) se encontram em um dos tipos de sentenças monorgumentais apresentadas por Castilho (2010). Para o autor, o verbo impessoal presente nessas estruturas foi gramaticalizado tornando-se funcional, o que explica a organização sintagmática *Verbo impessoal* + *SN*, sendo o SN o predicador.

Da mesma maneira que ocorre em outras construções, defendemos que o significado de uma estrutura ultrapassa os aspectos lexicais, por isso, em construções como a Existencial Apresentacional, o significado requer uma contribuição lexical, na medida que apenas verbos impessoais podem estar presentes nessa construção, e uma contribuição construcional, já que temos um SN com função apresentativa.

A modelagem da Construção Existencial Apresentacional considerou a existência de um núcleo verbal que deve ser preenchido por verbo existencial tendo como complemento um Sintagma Nominal, e ainda podendo haver um ou mais predicados secundários. A Figura 91 mostra a construção no Constructicon com suas principais características.

Figura 91 – Construção Existencial Apresentacional

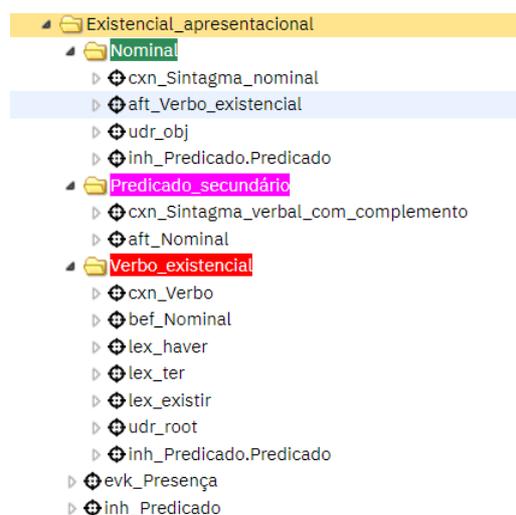
Existencial_apresentacional []	
Definição	
Um núcleo verbal preenchido por um verbo impessoal e tem como complemento um Sintagma Nominal, o Nominal . Seguindo o Nominal , pode haver um ou mais Predicado_secundário .	
Exemplo(s)	
Elementos da Construção	
Nominal [ent]	O Sintagma nominal que é apresentado ao receptor.
Predicado_secundário [sec]	Uma descrição do Nominal .
Verbo_existencial [the]	O núcleo verbal preenchido pelo verbo existencial.
Relações	
Evoca	Presença
Herda de	Predicado

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

O CE Nominal que é constituído de SN que na Construção Existencial Apresentacional é uma entidade, muitas vezes, aparecerá juntamente com uma expressão locativa como em (92) *ali* e (93) *na cozinha*.

Para modelar adequadamente essa construção a restrição CE>lexeme, uma vez que apenas os verbos *ter*, *haver* e *existir* podem preencher o *slot* do núcleo verbal na construção. Como podemos ver na figura 92.

Figura 92 – Restrições aplicadas aos ECs da Construção Existencial Apresentacional

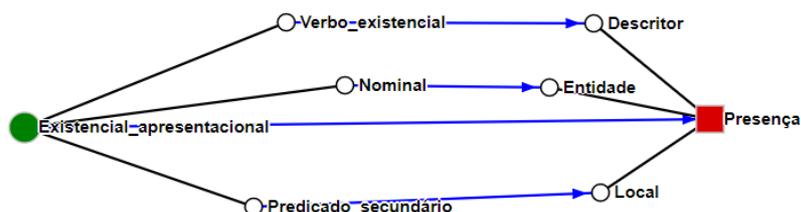


Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

O significado da Construção Existencial Apresentacional mapeado através do *frame* Presença em que uma Entidade existe em um determinado Local, em um determinado Tempo, conforme observado por um observador implícito.

A Figura 93 apresenta como foi realizado o mapeamento dos ECs da construção com o respectivo *frame* evocado por ela. É sempre bom ressaltar que mesmo a unificação ocorrendo entre ECs e EFs o significado é uma junção da semântica lexical com a semântica construcional, assim, o *frame* Presença através dos EFs indica como uma certa Entidade que é representada pelo EC Nominal se encontra presente em um local especificado pelo Predicado_secundário.

Figura 93 – Frame evocado pela Construção Existencial Apresentacional



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

A última construção da família sem sujeito é a Construção de Tempo Decorrido, que apresenta a mesma estrutura da Existencial Apresentacional [SV SN] com a diferença que a Construção de Tempo Decorrido apresenta a constatação de algo que ocorreu há determinado tempo, como em (94) e (95).

(94) Faz cinco anos que me casei.

(95) Há três anos mudei de cidade.

Castilho (2010) engloba construtos como (94) e (95) nas sentenças apresentacionais, afirmando que para algo existir requer uma complementação locativa ou temporal, sendo nesse caso a complementação temporal responsável pela predicação.

A Construção de Tempo Decorrido possui características semelhantes à Existencial Apresentacional um verbo impessoal, um SN que irá indicar o tempo e um predicado secundário responsável por trazer uma descrição ou especificação para o SN. A Figura 94 mostra como a construção foi modelada no Constructicon, seguindo essas características.

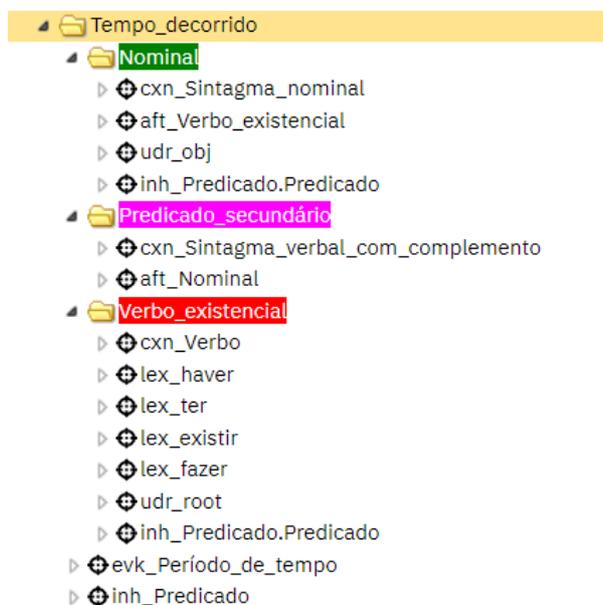
Figura 94 – Construção Tempo Decorrido

Tempo_decorrido [Tempo_decorrido]	
Definição	
Um núcleo verbal preenchido por um verbo existencial e tem como complemento um Sintagma Nominal, o Nominal . Seguindo o Nominal, pode haver um ou mais Predicado_secundário . Nesta construção temos a constatação de algo que ocorreu há determinado tempo	
Exemplo(s)	
Elementos da Construção	
Nominal [Nominal]	O Sintagma nominal indefinido que é representado por uma expressão temporal.
Predicado_secundário [Predicado_secundário]	Uma descrição do Nominal
Verbo_existencial [Verbo_existencial]	O núcleo verbal preenchido pelo verbo existencial
Relações	
Evoca	Período_de_tempo
Herda de	Predicado

Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Um ponto que difere a Construção de Tempo Decorrido da Existencial Apresentacional diz respeito aos verbos que devem preencher o *slot* do núcleo verbal, além dos verbos *ter*, *haver* e *existir*, na Construção de Tempo Decorrido ainda podemos incluir o verbo *fazer*. Essa informação foi registrada na modelagem da restrição CE>lexeme como podemos ver na Figura 95.

Figura 95 – Restrições aplicadas aos ECs da Construção Tempo Decorrido

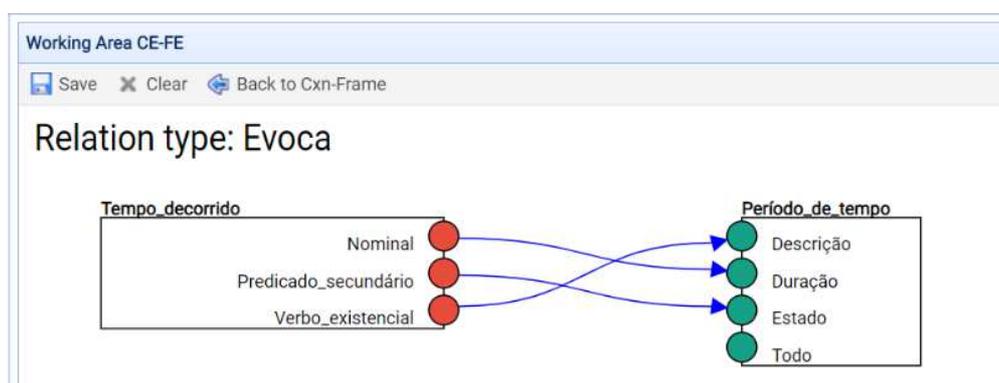


Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

Na Figura 96 encontramos como as características sintáticas e semânticas foram modeladas através das restrições e da unificação dos ECs com os EFs do *frame*

Período_de_tempo evocado pela construção, indicando a ocorrência de algo há um determinado tempo. Assim temos, o EC Nominal associado ao EF Duração e o EC Predicado_secundário associado ao EF Estado. Por exemplo, em construtos como (95) o SN *três anos* representa o EC Nominal que está correlacionado a caracterização do estado de duração, enquanto o Predicado_secundário representado por *mudei de cidade* refere-se a um estado em andamento a partir do qual uma duração pode ser medida.

Figura 96 – Relação Evoca Construção Tempo Decorrido



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

O nosso objetivo ao apresentar um modelo de representação linguístico-computacional para as construções de estrutura argumental no Português Brasileiro, além de contribuir para a constituição de uma gramática, propõe também uma sistematização de estruturas frequentes e já discutidas sob diferentes abordagens em nossa língua. O que propomos aqui é um modelo de representação do conhecimento linguístico, uma vez que esse envolve diversas e variadas dimensões, que tratam das questões fonéticas, morfológicas, sintáticas, semânticas e pragmáticas. Em especial, aborda a complexa questão do significado e de como ele pode ser representado.

As expressões linguísticas – motivadas pela cognição humana – são inerentemente flexíveis, ambíguas, dependentes do contexto, movidas pela criatividade e pouco aderentes a regras (FAUCONNIER & TURNER, 2002). Dessa forma, o que vimos durante muito tempo na busca de uma representação computacional mais adequada para o conhecimento linguístico, foi a negligência em relação a linguagem. A linguagem foi considerada autônoma e objetiva, e sua representação reduzida a regras lógico-formais. Os estudos na Linguística Cognitiva, no entanto, já mostraram a insuficiência do

significante para tratar da dimensão da significação linguística (SALOMÃO, 1999; FAUCONNIER & TURNER, 2002).

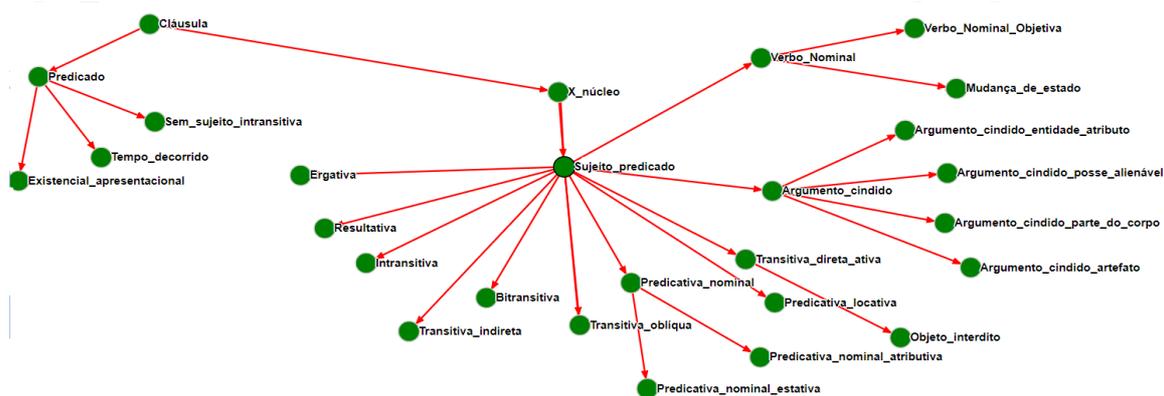
O problema não é representar o conhecimento através do uso sistemático da forma. O problema é considerar que essas representações abarcam todas as dimensões do significado. A formalização do conhecimento linguístico é necessária no âmbito computacional, mas não se deve ignorar a limitação que esta formalização impõe. Confiar que as formas são capazes de processar a essência dos sistemas conceituais de conhecimento é dizer que a capacidade humana de construir significados pode ser manipulada totalmente. No entanto, como afirmam Fauconnier e Turner (2002), as expressões linguísticas apontam para o significado, mas não o portam.

Esses limites à representação computacional do conhecimento podem ser amenizados por teorias linguísticas como a Semântica de Frames e algumas abordagens da Gramática de Construções, que buscam ser cognitivamente plausíveis. Como foi visto, a *FrameNet* e o *Constructicon* aliam o registro dos fenômenos linguísticos ao registro do significado associado a esses fenômenos. A rede de frames, elementos de frame, unidades lexicais e construções, representando o conhecimento linguístico, constitui uma ferramenta flexível e abrangente para o estudo e a compreensão da língua sob o ponto de vista computacional.

As construções que foram aqui modeladas explicitam a necessidade de considerar tanto aspectos semânticos como aspectos sintáticos para uma representação satisfatória do conhecimento linguístico. Para isso, de um lado, as restrições sintáticas foram tratadas em termos de sintagmas e categorias gramaticais como uma forma de mostrar as generalizações presentes em cada construção. Por outro lado, cada restrição foi unificada a um determinado componente semântico do *frame* evocado pela construção.

A proposta foi abarcar no *Constructicon* diferentes construções de estrutura argumental com o intuito de constituir uma gramática do PB que inclua diferentes padrões construcionais, já que esses manifestam *frames* genéricos das ações humanas básicas, que já estão disponíveis na base da *FrameNet* Brasil. Diante disso, uma vez essas construções modeladas em termos de restrições e relações de Herança e Evocação de *frames* é possível propor então uma rede de construções de estrutura argumental no Português Brasileiro como demonstra a Figura 97.

Figura 97 – Rede de construções de Estrutura Argumental



Fonte: Webtool (<http://webtool.framenetbr.ufjf.br/index.php/webtool/main>, 2020)

6 RECONHECIMENTO E VALIDAÇÃO DE CONSTRUÇÕES

Neste capítulo apresentaremos os resultados da aplicação do modelo de reconhecimento de construções descrito no capítulo 4, de metodologia. Como foi exposto no capítulo 3, encontramos diversos modelos de reconhecimento de construções tanto para o Português Brasileiro quanto para outras línguas. O que será discutido aqui é em que medida um modelo, que considerou as formalizações realizadas no *Constructicon*, fomentado pelas relações de UD, reconheceu as construções discutidas no capítulo anterior. O reconhecimento das construções pelo nosso modelo consistiu, primeiramente, na realização das anotações de instâncias das construções que foram modeladas. Posteriormente, essas mesmas sentenças anotadas foram submetidas à análise pelo modelo, para que realizasse a tarefa de reconhecimento. Nas próximas seções, apresentaremos os resultados das análises realizadas, como também a discussão sobre esses resultados.

6.1 RESULTADOS E PRIMEIRA DISCUSSÃO

A anotação das construções de estrutura argumental no *Corpus UD_Portuguese-Bosque* resultou em 748 instâncias, dentre as quais observamos em maior número as construções Transitiva Direta Ativa, Predicativa Nominal Atributiva e Transitiva Oblíqua. Na Tabela 7, temos as construções por número de instâncias anotadas.

Tabela 7 – Número de anotações por construção

Construção	Instâncias
Transitiva_direta_ativa	406
Predicativa_nominal_atributiva	108
Transitiva_oblíqua	75
Predicativa_nominal_estativa	37
Intransitiva	36
Ergativa	30
Bitransitiva_ativa	21
Predicativa_locativa	17
Existencial_apresentacional	9
Transitiva_indireta	7
Tempo_decorrido	2

Fonte: Criado pela autora

No capítulo anterior apresentamos a modelagem linguístico-computacional de 24 construções de Estrutura Argumental, no entanto, como podemos observar na tabela, nem todas foram instanciadas no *Corpus UD_Portuguese-Bosque*. Também como mencionado anteriormente, em paralelo à modelagem, a anotação em um *corpus* de treinamento teve o objetivo de identificar as construções modeladas e submeter essas anotações a um algoritmo capaz de reconhecer as construções anotadas.

A Tabela 7 demonstra que, das 24 construções modeladas, 11 apresentaram instâncias em corpus, a Transitiva_direta_ativa 406, a Predicativa_nominal_atributiva com 108, a Transitiva_oblíqua 75, a Predicativa_nominal_estativa com 37, a Intransitiva 36, a Ergativa com 30, Bitransitiva_ativa 21, a Predicativa_locativa com 17 Existencial_apresentacional com 9 instâncias, a Transitiva_indireta com 7 e a Tempo_decorrido com 2.

O *Corpus UD_Portuguese-Bosque* apresenta 3 *splits* que já se encontram anotados para a UD, uma vez que as sentenças presentes no *corpus* foram processadas pelo *parser* da UD e posteriormente anotadas por humanos. Em nosso experimento, escolhemos um dos *splits* randomicamente, já que o nosso intuito era que o modelo usasse a etiqueta da UD já atribuída como um recurso a mais para reconhecer construções, por isso, um *corpus* anotado por humanos com as etiquetas da UD pode melhorar o desempenho do reconhecedor de construções.

Nossa tarefa foi reconhecer instâncias das construções modeladas em um *corpus* anotado para dependências. Para esse fim, treinamos um modelo de IA e esse modelo foi até esse *corpus* anotado para identificar as construções e validá-las conforme a anotação. Como apresentamos no capítulo 4, o modelo foi treinado usando o BERT, que é um modelo de língua que não foi originalmente desenhado para reconhecer construções. O BERT é um modelo treinado para tarefas genéricas de Processamento de Língua Natural. A partir do BERT base, foram acrescentados novos modelos.

As Tabelas 8 e 9 apresentam como as 11 construções que receberam a anotação em *corpus* foram processadas pelo nosso primeiro modelo de reconhecimento de construções o BERT-Base. A nossa tarefa focou em identificar não só as construções, mas também os ECs dessas construções, uma vez que as informações da construção em si se encontram nos ECs via restrições que abarcam tanto características sintáticas como semânticas.

Tabela 8 – Resultados BERT-Base - Globais

Construções e ECs	Precision	Recall	F1	Precision -micro	Recall- micro	F1-micro	F1- weighted	Epochs
Global (Todos ECs)	0,1961	0,2354	0,2007	0,7527	0,7527	0,7527	0,5065	17,0

Fonte: criado pela autora

Tabela 9 – Resultados BERT-Base Construções e ECs

Construções e ECs	Precision	Recall	F1
Bitransitiva_ativa	0,0302	0,1562	0,0482
Bitransitiva_ativa.Predicado	0,0604	0,3123	0,0964
Bitransitiva_ativa.Sujeito	0,0000	0,0000	0,0000
Ergativa	0,0893	0,1401	0,0875
Ergativa.Predicado	0,0286	0,0600	0,0340
Ergativa.Sujeito	0,1500	0,2201	0,1411
Existencial_apresentacional	0,0000	0,0000	0,0000
Existencial_apresentacional.Nominal	0,0000	0,0000	0,0000
Existencial_apresentacional.Predicado_secundário	0,0000	0,0000	0,0000
Existencial_apresentacional.Verbo_existencial	0,0000	0,0000	0,0000
Intransitiva	0,0000	0,0000	0,0000
Intransitiva.Predicado	0,0000	0,0000	0,0000
Intransitiva.Sujeito	0,0000	0,0000	0,0000
Predicativa_locativa	0,2871	0,3398	0,2793
Predicativa_locativa.Predicado	0,5600	0,5796	0,5337
Predicativa_locativa.Sujeito	0,0143	0,1000	0,0250
Predicativa_nominal_atributiva	0,7262	0,7290	0,7189
Predicativa_nominal_atributiva.Predicado	0,6667	0,7776	0,7141
Predicativa_nominal_atributiva.Sujeito	0,7857	0,6804	0,7238
Predicativa_nominal_estativa	0,2532	0,3793	0,2777
Predicativa_nominal_estativa.Predicado	0,3833	0,5369	0,3993
Predicativa_nominal_estativa.Sujeito	0,1231	0,2217	0,1561
Tempo_decorrido	0,0000	0,0000	0,0000
Tempo_decorrido.Nominal	0,0000	0,0000	0,0000
Tempo_decorrido.Predicado_secundário	0,0000	0,0000	0,0000
Tempo_decorrido.Verbo_existencial	0,0000	0,0000	0,0000
Transitiva_direta_ativa	0,6519	0,6594	0,6540
Transitiva_direta_ativa.Predicado	0,5763	0,5688	0,5710
Transitiva_direta_ativa.Sujeito	0,7275	0,7500	0,7370
Transitiva_indireta	0,0000	0,0000	0,0000
Transitiva_indireta.Predicado	0,0000	0,0000	0,0000
Transitiva_indireta.Sujeito	0,0000	0,0000	0,0000
Transitiva_obliqua	0,3151	0,4209	0,3431
Transitiva_obliqua.Predicado	0,3413	0,4094	0,3619

Fonte: criado pela autora

A nossa tarefa de reconhecer construções foi testada usando as seguintes métricas que se encontram apresentadas na Tabela 8. As métricas mediram a acurácia do modelo, ou seja, a qualidade de reconhecer as construções a partir de diferentes fatores. É importante pontuar que a acurácia de cada métrica recebe um valor de 0 a 1, em que quanto mais próximo de 1 melhor é a qualidade do modelo.

As métricas que constam nas tabelas são:

- (i) **Precision** que é calculada quando temos todas as instâncias verdadeiramente positivas e esse número é dividido pela soma das instâncias que nós identificamos, juntamente com as que o modelo identificou. Assim temos, dentre os elementos que o modelo classificou como uma construção específica, quantos de fato são essa construção ou esse elemento da construção. Por exemplo, como mostra a Tabela 8, quando temos uma Precision micro para todos os ECs de 0,7527, isso significa que, em um universo de 100 construções, em cerca de 25 ele classificou elementos que não eram de uma construção específica como sendo de uma construção específica.
- (ii) **Recall** ou revocação é calculada com base no número de instâncias que nós identificamos e o modelo não viu, que chamamos de falso negativo. Nesse sentido, mede, dentre todas as construções ou ECs específicos que o modelo deveria ter identificado, quantos de fato ele identificou.
- (iii) **F1** é a métrica que combina Precision e Recall. Ela mapeia o quanto o modelo identificou, ou seja, o quanto de uma construção ou ECs específicos existentes no *corpus* ele identificou corretamente.
- (iv) As métricas **Precision- micro**, **Recall-micro**, **F1-micro** e **F1-weighted**, **Epochs** são globais, por isso, aparecem identificadas apenas quando consideramos todas as construções e todos os ECs. É o cálculo das métricas de forma geral para o conjunto das construções encontradas.

Os outros modelos testados buscaram medir de que maneira o acréscimo de novas informações poderia contribuir para a identificação de construções. Dessa forma, os próximos resultados são da incrementação do BERT-Base com as relações sintáticas da UD, juntamente com as relações de ordem dos constituintes na sentença.

As Tabela 10 e 11 mostram os resultados desse novo modelo. Se compararmos as métricas globais que consideram todas as construções e todos os ECs em relação às Tabelas 8 e 9, com os resultados do BERT-Base, vemos que houve uma mudança na acurácia. Enquanto o BERT-Base apresenta um F1 micro de 0,7527 e um F1 weighted de 0,5065, o modelo BERT-Base + UD (rel & ordem) apresenta um F1 micro de 0,7645 e um F1 1 weighted de 0,5424.

Tabela 10 - Resultados BERT-Base + UD (rel & ordem) - Globais

Construções e Ecs	Precision	Recall	F1	Precision-micro	Recall-micro	F1-micro	F1-weighted	Epochs
Global (Todos Ecs)	0,2480	0,2903	0,2518	0,7645	0,7645	0,7645	0,5424	22.2

Fonte: criado pela autora

Tabela 11 - Resultados BERT-Base + UD (rel & ordem) - Construções e ECs

Construções e Ecs	Precision	Recall	F1
Bitransitiva_ativa	0,0766	0,2605	0,1101
Bitransitiva_ativa.Predicado	0,1104	0,3328	0,1518
Bitransitiva_ativa.Sujeito	0,0429	0,1881	0,0684
Ergativa	0,2330	0,1817	0,1922
Ergativa.Predicado	0,2286	0,1613	0,1810
Ergativa.Sujeito	0,2375	0,2020	0,2034
Existencial_apresentacional	0,0053	0,0444	0,0094
Existencial_apresentacional.Nominal	0,0158	0,1333	0,0281
Existencial_apresentacional.Predicado_secundário	0,0000	0,0000	0,0000
Existencial_apresentacional.Verbo_existencial	0,0000	0,0000	0,0000
Intransitiva	0,0792	0,1129	0,0874
Intransitiva.Predicado	0,0333	0,0658	0,0415
Intransitiva.Sujeito	0,1250	0,1600	0,1333
Predicativa_locativa	0,3750	0,4155	0,3675
Predicativa_locativa.Predicado	0,6500	0,6596	0,6192
Predicativa_locativa.Sujeito	0,1000	0,1714	0,1159
Predicativa_nominal_tributiva	0,7558	0,7971	0,7709
Predicativa_nominal_tributiva.Predicado	0,7056	0,8308	0,7614
Predicativa_nominal_tributiva.Sujeito	0,8061	0,7634	0,7804
Predicativa_nominal_estativa	0,3808	0,4916	0,3892
Predicativa_nominal_estativa.Predicado	0,6000	0,5900	0,5603
Predicativa_nominal_estativa.Sujeito	0,1615	0,3932	0,2182
Tempo_decorrido	0,0000	0,0000	0,0000
Tempo_decorrido.Nominal	0,0000	0,0000	0,0000

Tempo_decorrido.Predicado_secundário	0,0000	0,0000	0,0000
Tempo_decorrido.Verbo_existencial	0,0000	0,0000	0,0000
Transitiva_direta_ativa	0,6625	0,6895	0,6740
Transitiva_direta_ativa.Predicado	0,5874	0,5989	0,5916
Transitiva_direta_ativa.Sujeito	0,7377	0,7800	0,7564
Transitiva_indireta	0,0000	0,0000	0,0000
Transitiva_indireta.Predicado	0,0000	0,0000	0,0000
Transitiva_indireta.Sujeito	0,0000	0,0000	0,0000
Transitiva_obliqua	0,4054	0,4686	0,4165
Transitiva_obliqua.Predicado	0,3663	0,4021	0,3778
Transitiva_obliqua.Sujeito	0,4444	0,5350	0,4551

Fonte: criado pela autora

As Tabelas 12, 13, 14 e 15. apresentam os resultados para os outros dois modelos testados com a informação acrescida dos *frames*.

Tabela 12 - Resultados BERT-Base + UD (rel & ordem) + Frame - Globais

Construções e Ecs	Precision	Recall	F1	Precision -micro	Recall- micro	F1-micro	F1- weighted	Epochs
Global (Todos Ecs)	0,2315	0,2580	0,2287	0,7581	0,7581	0,7581	0,5287	23,0

Fonte: criado pela autora

Tabela 13 - Resultados BERT-Base + UD (rel & ordem) + Frame - Construções e ECs

Construções e Ecs	Precision	Recall	F1
Bitransitiva_ativa	0,0769	0,2810	0,1110
Bitransitiva_ativa.Predicado	0,1396	0,3953	0,1962
Bitransitiva_ativa.Sujeito	0,0143	0,1667	0,0258
Ergativa	0,2674	0,2302	0,2315
Ergativa.Predicado	0,2286	0,2179	0,2047
Ergativa.Sujeito	0,3063	0,2425	0,2583
Existencial_apresentacional	0,0009	0,0167	0,0017
Existencial_apresentacional.Nominal	0,0026	0,0500	0,0050
Existencial_apresentacional.Predicado_secundário	0,0000	0,0000	0,0000
Existencial_apresentacional.Verbo_existencial	0,0000	0,0000	0,0000
Intransitiva	0,0125	0,0071	0,0091
Intransitiva.Predicado	0,0000	0,0000	0,0000
Intransitiva.Sujeito	0,0250	0,0143	0,0182
Predicativa_locativa	0,2521	0,3177	0,2550
Predicativa_locativa.Predicado	0,3900	0,4754	0,3766

Predicativa_locativa.Sujeito	0,1143	0,1600	0,1333
Predicativa_nominal_atributiva	0,7397	0,7523	0,7401
Predicativa_nominal_atributiva.Predicado	0,7222	0,7816	0,7487
Predicativa_nominal_atributiva.Sujeito	0,7571	0,7231	0,7316
Predicativa_nominal_estativa	0,4162	0,4227	0,3801
Predicativa_nominal_estativa.Predicado	0,6556	0,5302	0,5530
Predicativa_nominal_estativa.Sujeito	0,1769	0,3152	0,2072
Tempo_decorrido	0,0000	0,0000	0,0000
Tempo_decorrido.Nominal	0,0000	0,0000	0,0000
Tempo_decorrido.Predicado_secundário	0,0000	0,0000	0,0000
Tempo_decorrido.Verbo_existencial	0,0000	0,0000	0,0000
Transitiva_direta_ativa	0,6695	0,6753	0,6699
Transitiva_direta_ativa.Predicado	0,5960	0,5945	0,5937
Transitiva_direta_ativa.Sujeito	0,7430	0,7562	0,7460
Transitiva_indireta	0,0000	0,0000	0,0000
Transitiva_indireta.Predicado	0,0000	0,0000	0,0000
Transitiva_indireta.Sujeito	0,0000	0,0000	0,0000
Transitiva_obliqua	0,3420	0,3848	0,3457
Transitiva_obliqua.Predicado	0,3228	0,3807	0,3274
Transitiva_obliqua.Sujeito	0,3611	0,3888	0,3640

Fonte: criado pela autora

Tabela 14 - Resultados Frames - Globais

Construções e Ecs	Precision	Recall	F1	Precision-micro	Recall-micro	F1-micro	F1-weighted	Epochs
Global (Todos CEs)	0,2179	0,2494	0,2193	0,7505	0,7505	0,7505	0,5182	18,9

Fonte: criado pela autora

Tabela 15 - Resultados Frames – Construções e Ecs

Construções e Ecs	Precision	Recall	F1
Bitransitiva_ativa	0,0516	0,2326	0,0815
Bitransitiva_ativa.Predicado	0,0937	0,3651	0,1456
Bitransitiva_ativa.Sujeito	0,0095	0,1000	0,0174
Ergativa	0,2402	0,1857	0,1981
Ergativa.Predicado	0,1429	0,1011	0,1178
Ergativa.Sujeito	0,3375	0,2704	0,2784
Existencial_apresentacional	0,0000	0,0000	0,0000
Existencial_apresentacional.Nominal	0,0000	0,0000	0,0000
Existencial_apresentacional.Predicado_secundário	0,0000	0,0000	0,0000
Existencial_apresentacional.Verbo_existencial	0,0000	0,0000	0,0000

Intransitiva	0,0250	0,0163	0,0194
Intransitiva.Predicado	0,0000	0,0000	0,0000
Intransitiva.Sujeito	0,0500	0,0325	0,0389
Predicativa_locativa	0,2579	0,2990	0,2689
Predicativa_locativa.Predicado	0,4300	0,4980	0,4456
Predicativa_locativa.Sujeito	0,0857	0,1000	0,0923
Predicativa_nominal_atributiva	0,7600	0,7138	0,7252
Predicativa_nominal_atributiva.Predicado	0,6833	0,7743	0,7210
Predicativa_nominal_atributiva.Sujeito	0,8367	0,6533	0,7294
Predicativa_nominal_estativa	0,2387	0,3987	0,2743
Predicativa_nominal_estativa.Predicado	0,3389	0,4443	0,3534
Predicativa_nominal_estativa.Sujeito	0,1385	0,3531	0,1952
Tempo_decorrido	0,0000	0,0000	0,0000
Tempo_decorrido.Nominal	0,0000	0,0000	0,0000
Tempo_decorrido.Predicado_secundário	0,0000	0,0000	0,0000
Tempo_decorrido.Verbo_existencial	0,0000	0,0000	0,0000
Transitiva_direta_ativa	0,6636	0,6599	0,6574
Transitiva_direta_ativa.Predicado	0,5982	0,5536	0,5728
Transitiva_direta_ativa.Sujeito	0,7290	0,7662	0,7420
Transitiva_indireta	0,0000	0,0000	0,0000
Transitiva_indireta.Predicado	0,0000	0,0000	0,0000
Transitiva_indireta.Sujeito	0,0000	0,0000	0,0000
Transitiva_obliqua	0,3779	0,4866	0,4064
Transitiva_obliqua.Predicado	0,3891	0,3779	0,3744
Transitiva_obliqua.Sujeito	0,3667	0,5953	0,4385

Fonte: criado pela autora

Como vemos nas Tabelas 12, 13, 14 e 15, quando trazemos a informação dos frames para o modelo, há uma queda no desempenho global. Nas Tabelas 12 e 13 que consideram o BERT- Base + UD (rel & ordem) + Frame, além das informações de UD, o modelo considera também as informações dos *frames* evocados pelos itens lexicais das construções. Notamos que, enquanto o desempenho desse modelo melhora em relação ao BERT-Base, ele piora quando consideramos apenas o acréscimo das relações de UD. O modelo apresenta um F1 micro de 0,7581 e um F1 weighted de 0,5287.

Já o modelo apresentado nas Tabelas 14 e 15 – Frames – indica que nesse cenário temos os *frames* dos itens lexicais juntamente com as *frame families*, ou seja, se analisa a família dos *frames* via relação de herança. Os resultados desse modelo apontam para uma queda no desempenho global em relação ao BERT-Base e aos outros modelos, nesse caso temos um F1 micro de 0,7505 e um F1 weighted de 0,5182.

Ao analisarmos os resultados até aqui, temos que, primeiramente, algumas construções que foram modeladas não apresentaram nenhuma instância para ser anotada no *corpus*. Acreditamos que isso se deve ao fato de, por mais produtivas que essas construções sejam no Português Brasileiro, apenas um *corpus* com um número muito grande de sentenças e uma variedade maior de gêneros textuais abarcaria tais estruturas. Sardinha (2000) afirma que o *corpus* possui uma função representativa, assim, algumas estruturas têm frequência de ocorrência muito rara e, para que haja probabilidade de ocorrerem no *corpus*, é necessário incorporar-se uma quantidade grande de palavras ao *corpus*. Em outras palavras, o autor afirma que quanto maior a quantidade de palavras, mais probabilidade há de palavras de baixa frequência aparecerem.

O autor também pontua que palavras e estruturas com diferentes significados e configurações sofrem uma frequência variada. Assim, mesmo estruturas de alta frequência têm sentidos e configurações distintos, por exemplo, o verbo *beber* tem mais probabilidade de aparecer em uma construção Transitiva Direta Ativa do que na construção de Objeto Interdito, portanto, o *beber* do segundo caso terá maior probabilidade de ocorrer quanto maior for o *corpus*.

Assim, quando se diz que um *corpus* deva ser representativo, entende-se representatividade em termos da extensão do *corpus*, isto é, de um número determinado de palavras e de textos.

Como para nossa tarefa era necessário um *corpus* previamente anotado por humanos – uma vez que queríamos testar um modelo em que pudéssemos usar não só as informações da nossa anotação no *Constructicon*, também as relações de UD com o objetivo de enriquecer o modelo para o reconhecimento de construções –, esse fator acabou por representar uma limitação. Encontrar um *corpus* previamente anotado e processado como o *UD_Portuguese-Bosque* para o PB é uma tarefa um tanto complicada. Primeiro porque os *parsers* para o Português Brasileiro apresentam desempenho pouco satisfatório (ALMEIDA, 2016) e segundo porque um *corpus* anotado por humanos com os metadados da UD também não é fácil de ser encontrado. Dessa forma, o *UD_Portuguese-Bosque* se apresenta como pouco representativo e esse fator se revela no resultado dos modelos testados.

Notamos que, das 24 construções, apenas 11 receberam anotações e, por isso, foram submetidas aos testes de reconhecimento. Dessas 11 construções, apenas 2 delas apresentaram um número maior que 100 ocorrências: a construção Transitiva Direta Ativa e construção Predicativa Nominal Atributiva.

Por isso, quando analisamos os resultados globais, como o da métrica F1 micro, que mapeia o quanto o modelo cobre, interpretamos que os resultados são satisfatórios, porém a diferença no número de instâncias das construções anotadas nos indica que quando se tem muitas construções como a Transitiva Direta Ativa, que apresenta 406 anotações, o modelo aprende relativamente bem e acaba subindo o F1 micro por ser a classe mais representativa, mas isso não se reflete quando analisamos o F1 macro de cada construção ou CE porque esse resultado é muito baixo, uma vez que há construções que o modelo sequer aprendeu.

Diante dessas indagações vimos que, numa tarefa de reconhecimento de construções em que testamos quatro modelos com especificações diferentes, foi possível observar que, quando usamos apenas o BERT-Base com o acréscimo das relações de UD, houve uma melhora no desempenho, ou seja, mais construções foram reconhecidas, porém os modelos que acrescentam as informações relacionadas aos *frames* dos itens lexicais mais as famílias de *frames* desses itens não apresentaram melhora em relação ao modelo que usa apenas as relações da UD.

Esse fator nos levou a considerar que, talvez, se testássemos modelos que considerassem apenas as duas construções mais frequentes no *corpus* o desempenho poderia ser melhor. Assim testamos os mesmos modelos para a construção Transitiva Direta Ativa e Predicativa Nominal Atributiva.

Tabela 16 -Resultados BERT-Base para as construções Transitiva Direta Ativa e Predicativa Nominal Atributiva

Construções e CEs	Precision	Recall	F1	Precision -micro	Recall -micro	F1-micro	F1-weighted	Epochs
Global (Todos CEs)	0,6742	0,7295	0,6911	0,8273	0,8273	0,8273	0,6903	10,4
Predicativa_nominal_atributiva	0,6829	0,7221	0,6886					
Predicativa_nominal_atributiva. Predicado	0,6941	0,7311	0,7012					
Predicativa_nominal_atributiva. Sujeito	0,6717	0,7131	0,6760					
Transitiva_direta_ativa	0,6655	0,7369	0,6936					
Transitiva_direta_ativa. Predicado	0,6956	0,6705	0,6802					
Transitiva_direta_ativa.Sujeito	0,6354	0,8033	0,7069					

Fonte: criado pela autora

A Tabela 16 apresenta os resultados do BERT-Base para as duas construções com mais de 100 anotações. Quando observamos os números percebemos uma melhora no

modelo que usa apenas as informações do BERT. Se antes tínhamos uma F1 micro 0,7527 e uma F1 weighted de 0,5065, aqui temos os valores de 0,8273 e 0,6903 para as duas métricas citadas respectivamente.

Essa informação nos leva a crer que se o BERT-Base sozinho melhora o desempenho para o reconhecimento de construções, isso indica que ele está aprendendo com mais eficiência os padrões das construções. Porém, quando investigamos os resultados para os outros modelos que acrescentam as informações da UD e dos *frames*, percebemos uma queda de desempenho para a identificação de construções. Como demonstra as Tabelas 17, 18 e 19.

Tabela 17 - Resultados BERT-Base + UD (rel & ordem) para as construções Transitiva Direta Ativa e Predicativa Nominal Atributiva

Construções e ECs	Precision	Recall	F1	Precision -micro	Recall -micro	F1-micro	F1-weighted	Epochs
Global (Todos CEs)	0,6121	0,7111	0,6504	0,8219	0,8219	0,8219	0,6751	10
Predicativa_nominal_atributiva	0,5525	0,6874	0,6040					
Predicativa_nominal_atributiva .Predicado	0,6050	0,6828	0,6373					
Predicativa_nominal_atributiva .Sujeito	0,5000	0,6920	0,5707					
Transitiva_direta_ativa	0,6716	0,7349	0,6968					
Transitiva_direta_ativa. Predicado	0,6923	0,6702	0,6796					
Transitiva_direta_ativa.Sujeito	0,6510	0,7996	0,7141					

Fonte: criado pela autora

Tabela 18 - BERT-Base + UD (rel & ordem) + Frame para as construções Transitiva Direta Ativa e Predicativa Nominal Atributiva

Construções e ECs	Precision	Recall	F1	Precision -micro	Recall -micro	F1-micro	F1-weighted	Epochs
Global (Todos CEs)	0,5930	0,6808	0,6173	0,8192	0,8192	0,8192	0,6686	10,4
Predicativa_nominal_atributiva	0,5035	0,6292	0,5340					
Predicativa_nominal_atributiva .Predicado	0,6465	0,6262	0,6324					
Predicativa_nominal_atributiva .Sujeito	0,3604	0,6322	0,4357					
Transitiva_direta_ativa	0,6825	0,7324	0,7006					
Transitiva_direta_ativa. Predicado	0,7110	0,6746	0,6890					
Transitiva_direta_ativa.Sujeito	0,6540	0,7902	0,7122					

Fonte: criado pela autora

Tabela 19 - Resultados Frames para as construções Transitiva Direta Ativa e Predicativa Nominal Atributiva

Construções e ECs	Precision	Recall	F1	Precision -micro	Recall- micro	F1-micro	F1- weighted	Epochs
Global (Todos CEs)	0,5982	0,7215	0,6388	0,8231	0,8231	0,8231	0,6666	10,2
Predicativa_nominal_atributiva	0,5506	0,6905	0,5871					
Predicativa_nominal_atributiva .Predicado	0,6446	0,6867	0,6593					
Predicativa_nominal_atributiva .Sujeito	0,4566	0,6943	0,5150					
Transitiva_direta_ativa	0,6458	0,7525	0,6905					
Transitiva_direta_ativa. Predicado	0,6472	0,6834	0,6625					
Transitiva_direta_ativa.Sujeito	0,6444	0,8216	0,7185					

Fonte: criado pela autora

As Tabelas 17, 18 e 19 apresentam os resultados dos modelos BERT-Base + UD (rel & ordem), BERT-Base + UD (rel & ordem) + Frame e do modelo que considera os *frames* dos itens lexicais mais a família de frames a qual esses itens estão relacionados. Enquanto tínhamos uma F1 micro de 0,8273 no BERT-Base, os modelos com as novas informações apresentam uma F1 micro de 0,8219 para quando adicionamos as informações das relações sintáticas e de ordem de constituintes da UD, uma F1 micro de 0,8192 quando além das informações da UD adicionamos a informação dos *frames* conforme os itens lexicais e uma F1 micro de 0,8231 quando consideramos apenas as informações de *frames* e suas cadeias de herança.

Como podemos perceber, os resultados apontam para uma queda no desempenho quando adicionamos novas informações ao modelo base para reconhecer as duas construções com maior número de ocorrências. Quando comparamos os modelos que reconheceram primeiramente as 11 construções e posteriormente apenas duas, notamos que, ao acrescentarmos as informações da UD nos testes com as 11 construções, há um aumento no desempenho diferentemente do que acontece quando testamos apenas duas.

Ao ressaltar essa informação chegamos à conclusão de que, se temos uma estrutura com muitas anotações, como é o caso das construções Transitiva Direta Ativa e Predicativa Nominal Atributiva, a informação estrutural, que se acrescenta via *constraints* faz pouca diferença. Nesses casos, o esforço de modelagem quando se tem um padrão muito frequente não trouxe muitas contribuições.

A estrutura dessas construções, quando as analisamos separadamente, é muito diferente. Enquanto na construção Predicativa Nominal Atributiva temos a configuração

de um Predicado constituído por verbo cópula mais um adjetivo, na construção Transitiva Direta Ativa temos um predicador verbal prototípico e não cópulas. Essa informação pode ser observada no desempenho para o reconhecimento dos CEs. O Predicado da Transitiva Direta Ativa tem um desempenho pior no modelo base que acrescenta relações de UD em que o F1 micro é de 0,6802, porém no modelo em que além das relações de UD temos o acréscimo do frame o F1 micro é de 0,6890

Esses resultados nos fizeram questionar qual seria o ganho real da modelagem de construções para tarefas de reconhecimento. Então nos atentamos para outro dado que aparece nos resultados. Se, por um lado, a modelagem pouco contribui para construções com mais anotações no corpus, por outro, quando olhamos para outras construções através dos ECs, vemos que, se a construção não apresentar um padrão construcional muito frequente no corpus, a indicação é que modelar melhora muito o desempenho. Na próxima seção, apresentaremos esse outro olhar para os dados, pontuando em quais casos a modelagem proporcionou uma melhora no desempenho da tarefa de reconhecer construções.

6.2 UMA NOVA ANÁLISE

Os resultados apresentados e discutidos nas seções anteriores nos atentaram para outro fator que aparece nos dados: como de fato a modelagem pode contribuir para o reconhecimento de construções pouco frequentes no *corpus*. Vamos pontuar aqui as construções que foram reconhecidas de forma mais eficiente graças à modelagem. A Tabela 20 apresenta os resultados para a tarefa na construção Bitransitiva_ativa.

Tabela 20 - Resultados para a construção Bitransitiva_ativa

Modelos	Precision	Recall	F1
BERT-Base	0,0302	0,1562	0,0482
BERT-Base + UD (rel & ordem)	0,0766	0,2605	0,1101
BERT-Base + UD (rel & ordem) + Frame	0,0769	0,2810	0,1110
Frames	0,0516	0,2326	0,0815

Fonte: criado pela autora

O que esses resultados nos informam é que o reconhecimento da construção Bitransitiva ativa, que apresentava um F1 micro no BERT-Base de 0,0482, quando considera as informações da UD juntamente com as informações do *frame* dos itens lexicais, melhora seu desempenho na tarefa, já que a configuração que usa essas novas informações apresenta um resultado de 0,1110, indicando que o modelo reconheceu melhor essa estrutura.

A construção *Bitransitiva_ativa*, que contou com 21 instâncias anotadas no *corpus*, é a única construção triargumental em nossa modelagem e é organizada de modo que temos um argumento externo e dois argumentos internos. A modelagem apontou que esses argumentos internos constituem o Predicado através de um sintagma verbal bitransitivo e o argumento externo é um sintagma nominal que constitui o EC Sujeito.

Ao analisarmos os resultados dos modelos para o reconhecimento dos ECs da construção *Bitransitiva_ativa*, percebemos como, a partir das informações das relações de UD e dos *frames* acrescentadas ao modelo base, os ECs foram mais bem reconhecidos, como demonstra a Tabela 21.

Tabela 21 - Resultados para ECs da construção *Bitransitiva_ativa*

Modelos	Bert-Base		Bert + UD		Bert + UD + Frame		Frames	
	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado
Precision	0,0000	0,0604	0,0429	0,1104	0,0143	0,1396	0,0095	0,0937
Recall	0,0000	0,3123	0,1881	0,3328	0,1667	0,3953	0,1000	0,3651
F1	0,0000	0,0964	0,0684	0,1518	0,0258	0,1962	0,0174	0,1456

Fonte: criado pela autora

Na Tabela 21, podemos ver que o Sujeito não era reconhecido pelo BERT-Base. Quando adicionamos outras informações, ele passa a ser reconhecido, tendo seu melhor resultado de 0,0684 na F1 no modelo que une o BERT-Base as informações de relação da UD juntamente com a ordem dos constituintes. Vemos que, na análise geral, o reconhecimento do EC-Sujeito apresenta um desempenho melhor em todos os outros modelos em relação ao BERT-Base. Isso acontece da mesma forma para o EC-Predicado.

Portanto, o que esses resultados apontam é que a modelagem pode ser muito útil para construções pouco frequentes em *corpus*, uma vez que, além de trazer novas informações sobre a estrutura das construções, ela é também responsável por validar aquelas informações que já foram apontadas.

Outro desempenho de reconhecimento que pode ser observado é o da construção Ergativa. A construção Ergativa consiste em uma construção de caráter não causativo, ou seja, temos um evento que afeta uma entidade. Nessa construção, temos um Predicado composto por um sintagma verbal sem completo e um Sujeito composto por um sintagma nominal.

Ao observarmos os resultados dos modelos para o reconhecimento da construção Ergativa, como mostra a Tabela 22, vemos que a construção é mais bem reconhecida quando comparamos o desempenho do BERT-Base em relação aos outros modelos. Com o BERT-Base temos uma F1 de 0,0875 que salta para 0,2315 no modelo que considera BERT-Base junto com as relações da UD (rel & ordem) e os *frames* dos itens lexicais da construção.

Tabela 22 - Resultados para a construção Ergativa

Modelos	Precision	Recall	F1
BERT-Base	0,0893	0,1401	0,0875
BERT-Base + UD (rel & ordem)	0,2330	0,1881	0,1922
BERT-Base + UD (rel & ordem) + Frame	0,2674	0,2302	0,2315
Frames	0,2402	0,1857	0,1981

Fonte: criado pela autora

Quando nos atentamos para os resultados dos ECs da construção Ergativa, percebemos resultados semelhantes àqueles da construção geral, ou seja, uma melhora no desempenho de reconhecimento dos outros modelos em relação ao BERT-Base, que apresenta uma F1 de 0,1411 para o Sujeito e 0,0340 para o Predicado. No modelo que alia as informações da UD juntamente com os *frames* a F1 é 0,2583 para o Sujeito e 0,2047 para o Predicado conforme a Tabela 23.

Tabela 23 - Resultados para os ECs da construção Ergativa

Modelos	Bert-Base		Bert + UD		Bert + UD + Frame		Frames	
	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado
Precision	0,1500	0,0286	0,2375	0,2286	0,3063	0,2886	0,1429	0,3375
Recall	0,2201	0,6000	0,2020	0,1613	0,2425	0,2179	0,1011	0,2704
F1	0,1411	0,0340	0,2034	0,1810	0,2583	0,2047	0,1178	0,2784

Fonte: criado pela autora

Os resultados das construções Bitransitiva_ativa e Ergativa apontam como os esforços de modelagem podem contribuir para o reconhecimento automático de construções. Quando analisamos os resultados do BERT-Base, que é um modelo pré-treinado a partir de textos sem anotação, e comparamos aos outros modelos, percebemos que, diferentemente do que observamos nas construções Transitiva Direta Ativa e Predicativa Nominal Atributiva, em que a adição de novas informações piora o desempenho, para as construções menos frequentes tal adição traz uma melhora, no caso, da construção Ergativa percebemos que um melhor reconhecimento no modelo BERT-Base + UD (rel & ordem) + Frame indica que os frames são fundamentais nessa construção.

No entanto, observamos que, das construções que apresentaram alguma instância anotada, nem todas foram reconhecidas pelos modelos, como é o caso das construções Tempo Decorrido com 2 instâncias, Existencial Apresentacional com 9 instâncias e Transitiva Indireta com 7 instâncias anotadas. Essas construções, apesar de terem sido encontradas no *corpus*, não foram reconhecidas, indicando que poucas anotações não são suficientes para essa tarefa.

Há, porém, construções que, mesmo sendo pouco frequentes em comparação à Transitiva Direta Ativa e a Predicativa Nominal Atributiva, apresentaram um número em que se pôde observar o desempenho dos modelos para o reconhecimento dessas construções. Além das construções Bitransitiva_ativa e Ergativa, outras merecem um olhar mais atento sobre como foram reconhecidas, esse é o caso da construção Intransitiva.

A estrutura da construção Intransitiva é semelhante à Ergativa: é uma construção composta por um sintagma verbal sem complemento, que é o Predicado, e um sintagma nominal que é o Sujeito, porém a construção Intransitiva apresenta sim um caráter causativo evocando por exemplo o frame de Agir Intencionalmente.

A Tabela 24 mostra os resultados da construção Intransitiva que apresentou 36 instâncias anotadas no *corpus*. O BERT-Base não reconheceu a construção, mas ela foi reconhecida pelos outros modelos mesmo com um desempenho baixo e apresentando melhor resultado para o modelo que une o BERT mais as relações da UD com F1 de 0,0874.

Tabela 24 - Resultados para a construção Intransitiva

Modelos	Precision	Recall	F1
BERT-Base	0,0000	0,0000	0,0000
BERT-Base + UD (rel & ordem)	0,0792	0,1129	0,0874
BERT-Base + UD (rel & ordem) + Frame	0,0215	0,0071	0,0091
Frames	0,0250	0,0163	0,0194

Fonte: criado pela autora

Quando observamos o resultado da tarefa para os ECs da construção Intransitiva, percebemos que, da mesma maneira que aconteceu para a construção Bitransitiva, o BERT-Base também não reconheceu os ECs da construção. Nos outros modelos ao menos o Sujeito foi reconhecido apontado uma pequena melhora de desempenho graças às novas informações que foram adicionadas, como podemos ver na Tabela 25.

Portanto, mesmo o Predicado não sendo reconhecido nos modelos BERT-Base + UD (rel & ordem) + Frame e Frames, o reconhecimento do Sujeito já é melhor do que no modelo que usa apenas o BERT.

Tabela 25 - Resultados para os ECs da construção Intransitiva

Modelos	Bert-Base		Bert + UD		Bert + UD + Frame		Frames	
	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado
Precision	0,0000	0,0000	0,1250	0,0333	0,0250	0,0000	0,0500	0,0000
Recall	0,0000	0,0000	0,1600	0,0658	0,0143	0,0000	0,0325	0,0000
F1	0,0000	0,0000	0,1333	0,0415	0,0782	0,0000	0,0389	0,0000

Fonte: criado pela autora

Outra construção cujos resultados podemos observar é a construção Predicativa_locativa, com 17 instâncias anotadas no *corpus*. Essa construção indica a localização de uma entidade em um determinado espaço através de uma cópula e um núcleo adverbial. Como podemos observar na Tabela 26, os resultados para o reconhecimento da construção Predicativa_locativa mostram que a melhor F1 de 0,3675 se encontra no modelo que possui o BERT-Base juntamente com as informações da UD (rel & ordem), ou seja, quando as informações de caráter sintático e ordem dos

constituintes foram adicionadas ao modelo base houve um melhor reconhecimento da construção.

Por outro lado, nos modelos em que, para além das informações da UD, consideram-se também quais *frames* estão relacionados aos itens lexicais da construção, notamos que o desempenho da tarefa é inferior se comparamos até mesmo com o BERT-Base.

Tabela 26 - Resultados para a construção Predicativa_locativa

Modelos	Precision	Recall	F1
BERT-Base	0,2871	0,3398	0,2793
BERT-Base + UD (rel & ordem)	0,3750	0,4155	0,3675
BERT-Base + UD (rel & ordem) + Frame	0,2521	0,3177	0,2550
Frames	0,2579	0,2990	0,2689

Fonte: criado pela autora

Assim, temos que as informações trazidas da UD para o modelo foram mais significativas do que as informações sobre os *frames*, uma vez que o modelo BERT-Base + UD (rel & ordem) + Frame apresentou uma F1 de 0,2550 e o modelo Frames apresentou uma F1 de 0,2689 inferior a F1 do modelo base que foi de 0,2793.

Ao analisarmos os resultados para o reconhecimento dos ECs da construção Predicativa_locativa, conforme a Tabela 27, notamos um comportamento um pouco diferente dos resultados para a construção no geral. Enquanto a F1 do modelo que alia o BERT mais as informações da UD é de 0,1154 para o Sujeito e de 0,6192 para o Predicado o modelo Bert + UD + Frame apresentou uma F1 de 0,1333 para o Sujeito e 0,3766 para o Predicado, assim houve uma melhora de desempenho para reconhecer o Sujeito em relação tanto ao modelo base como para o modelo com informações sintáticas e de ordem dos constituintes trazidos pela UD. No entanto, para o Predicado, notamos que o modelo Bert + UD + Frame apresenta um desempenho inferior quando comparamos a esses dois modelos. O mesmo acontece para o modelo que usa apenas as informações dos *frames*, para o Sujeito temos uma F1 de 0,0923 que é melhor do que a F1 só com BERT e para o Predicado temos uma F1 de 0,4456 que é inferior aos modelos Bert-Base e Bert + UD.

Dessa forma, vemos que no caso dos ECs da Predicativa_locativa a modelagem colaborou para a identificação do Sujeito enquanto as relações da UD foram mais significativas para o reconhecimento do Predicado.

Tabela 27 - Resultados para os ECs da construção Predicativa_locativa

Modelos	Bert-Base		Bert + UD		Bert + UD + Frame		Frames	
	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado
Precision	0,0143	0,5600	0,1000	0,6500	0,1143	0,3900	0,0857	0,4300
Recall	0,1000	0,5796	0,1714	0,6596	0,1600	0,4754	0,1000	0,4980
F1	0,0250	0,5337	0,1154	0,6192	0,1333	0,3766	0,0923	0,4456

Fonte: criado pela autora

Quando consideramos as 11 construções que apresentaram instâncias no *corpus*, podemos observar a construção Predicativa_nominal_estativa, que apresenta um número de 37 anotações. Da mesma forma que a construção Predicativa_nominal_atributiva, a nominal estativa apresenta a configuração sintática de um Sujeito que é um sintagma nominal e um Predicado composto por uma cópula e um predicativo do sujeito, sendo este um núcleo adjetival que informa o estado do Sujeito.

Os resultados para o reconhecimento dessa construção estão apresentados na Tabela 28.

Tabela 28 - Resultados para a construção Predicativa_estativa

Modelos	Precision	Recall	F1
BERT-Base	0,2532	0,3793	0,2777
BERT-Base + UD (rel & ordem)	0,3808	0,4916	0,3892
BERT-Base + UD (rel & ordem) + Frame	0,4162	0,4227	0,3801
Frames	0,2387	0,3987	0,2743

Fonte: criado pela autora

O melhor desempenho da tarefa se deu no modelo que faz a junção entre BERT com as informações da UD. Se no modelo base tínhamos uma F1 de 0,2777, no modelo BERT-Base + UD (rel & ordem) temos uma F1 de 0,3892 indicando uma melhora no reconhecimento da construção. Já quando analisamos os resultados para o modelo que além das relações da UD acrescenta também as informações dos *frames* vemos uma melhora em relação do BERT-Base, mas um desempenho um pouco inferior ao modelo BERT-Base + UD (rel & ordem), uma vez que a F1 é de 0,3801 indicando que esse modelo é melhor para reconhecer a construção Predicativa_estativa do que o modelo com

apenas as informações do BERT. Por fim, o modelo que usa apenas as informações dos Frames apresenta um resultado um pouco inferior ao do modelo base com uma F1 de 0,2743, ou seja, o modelo que usa apenas o BERT é melhor para reconhecer a construção *Predicativa_estativa* do que o modelo com os *frames*.

Os resultados para o reconhecimento dos ECs da construção *Predicativa_estativa*, como mostra a Tabela 29 apresentam resultados semelhantes dos da construção em si. O melhor desempenho de reconhecimento se encontra no modelo Bert + UD, em que observamos uma F1 de 0,2182 para o Sujeito e 0,5603 para o Predicado. Com exceção do Predicado no modelo que considera apenas as informações dos Frames, os outros resultados apontam que houve uma melhora para reconhecer os ECs em todos os modelos em relação ao BERT-Base, isto é, os ECs foram mais bem reconhecidos quando foram adicionadas novas informações ao modelo base.

Tabela 29 - Resultados para os ECs da construção *Predicativa_estativa*

Modelos	Bert-Base		Bert + UD		Bert + UD + Frame		Frames	
	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado
Precision	0,1231	0,3833	0,1615	0,6000	0,1769	0,6556	0,1385	0,3389
Recall	0,2217	0,5369	0,3832	0,5900	0,3152	0,5302	0,3531	0,4443
F1	0,1561	0,3993	0,2182	0,5603	0,2072	0,5530	0,1952	0,3534

Fonte: criado pela autora

A última construção a ser analisada nessa seção é a construção *Transitiva_obliqua*, que foi a terceira construção mais frequente no *corpus*, com 75 instâncias. Essa construção, como vimos no capítulo anterior, é composta por um argumento externo e um argumento oblíquo. A estrutura sintagmática e funcional dessa construção é um sujeito que é um sintagma nominal e o Predicado é a junção de um verbo e do argumento oblíquo correspondente a um sintagma preposicional. Dada a sua diversidade semântica, essa construção evoca o *frame* de Evento.

Os resultados da tarefa de reconhecimento da construção *Transitiva_obliqua* se encontram na Tabela 30. Como pode ser observado, o melhor desempenho se encontra no modelo BERT-Base + UD (rel & ordem) com uma F1 de 0,4165, porém tanto o modelo BERT-Base + UD (rel & ordem) + Frame e o modelo Frames apresentam uma melhora

em relação ao modelo base, já que temos uma F1 de 0,3457 para o primeiro e 0,4064 para o segundo.

Pelo fato de a construção Transitiva_oblíqua evocar o *frame* de Evento, que é bastante genérico e está relacionado a vários *frames* e unidades lexicais, acreditamos que esse possa ser motivo do modelo que considera apenas os *frames* apresentar um desempenho melhor do que o modelo BERT-Base + UD (rel & ordem). Com isso, temos que as informações dos frames e as informações da UD que foram adicionados ao BERT melhoram o reconhecimento da construção Transitiva_oblíqua.

Tabela 30 - Resultados para a construção Transitiva_oblíqua

Modelos	Precision	Recall	F1
BERT-Base	0,3151	0,4209	0,3431
BERT-Base + UD (rel & ordem)	0,4054	0,4686	0,4165
BERT-Base + UD (rel & ordem) + Frame	0,3420	0,3848	0,3457
Frames	0,3779	0,4866	0,4064

Fonte: criado pela autora

Os ECs da construção Transitiva_oblíqua, ao serem submetidos à tarefa de reconhecimento, apresentaram os resultados presentes na Tabela 31. O modelo que apresentou melhor desempenho foi o Bert + UD com uma F1 de 0,4551 para o Sujeito e 0,3778 para o Predicado. O modelo que, para além das informações da UD, adiciona as informações do frame, apresenta um resultado melhor para o Sujeito em relação ao modelo base com uma F1 de 0,3640, mas o resultado do Predicado é inferior com uma F1 de 0,3257. Da mesma maneira que ocorreu na construção geral, o modelo só com os *frames* apresentou um resultado superior os BERT-Base e ao Bert + UD + Frame com uma F1 de 0,4385 para o Sujeito e 0,3744 para o Predicado indicando que as informações desse modelo são mais eficientes para reconhecer os ECs da Transitiva_oblíqua do que o modelo que além do frame traz o BERT juntamente com a UD

Tabela 31 - Resultados para os ECs da construção Transitiva_oblíqua

Modelos	Bert-Base		Bert + UD		Bert + UD + Frame		Frames	
	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado	Sujeito	Predicado
Precision	0,2889	0,3413	0,4444	0,3663	0,3611	0,3228	0,3667	0,3891
Recall	0,4325	0,4094	0,5350	0,4021	0,3888	0,3807	0,5953	0,3779
F1	0,3242	0,3431	0,4551	0,3778	0,3640	0,3257	0,4385	0,3744

Fonte: criado pela autora

6.3 CONCLUINDO A ANÁLISE

Na seção anterior, apresentamos como um olhar diferente para os resultados pode indicar como a modelagem pode contribuir em alguns casos para a identificação das construções. As 6 construções analisadas foram as que apresentaram algum resultado de reconhecimento, além das construções *Transitiva_direta_ativa* e *Predicativa_nominal_atributiva*, já que as construções *Existencial_apresentacional*, *Tempo Decorrido* e *Transitiva_indireta*, que apresentaram o menor número de anotações no *corpus*, não apresentaram resultados para o reconhecimento da construção e nem dos ECs.

Vimos que a modelagem realizada no *Constructicon* pode auxiliar em tarefas de reconhecimento de construções menos frequentes nos dados. Os resultados obtidos apontam para a importância da interação entre os recursos disponíveis na FN-Br. Dentre as 6 construções analisadas, as construções *Bitransitiva_ativa* e *Ergativa* e seus ECs apresentaram um melhor desempenho para o modelo que considera o BERT-Base juntamente com as informações da UD (rel & ordem) e dos *frames* relacionados à construção indicando como a modelagem pode ser útil para auxiliar as tarefas de reconhecimento de construções. As construções *Predicativa_nominal_estativa* e *Transitiva_obliqua* e seus ECs apresentaram melhor resultado para o modelo do BERT que alia as informações da UD, porém os resultados do modelo que acrescenta as informações dos *frames* são mais satisfatórios do que os resultados com apenas o modelo base. Já as construções *Intransitiva* e *Predicativa_locativa* apresentaram melhor desempenho no modelo BERT + UD, no entanto, os outros modelos apresentaram resultados inferiores ao modelo base com exceção dos resultados para alguns ECs.

Diante disso, entendemos que, de forma geral, assim como as informações sintáticas e de ordem de constituintes trazidas da UD, a modelagem aqui proposta, que fornece não apenas informações sintáticas, mas também semânticas das construções, consegue contribuir para que construções possam ser reconhecidas em tarefas como também proporciona informações para que essas estruturas possam ser aprendidas por um algoritmo de IA, ainda que poucas instâncias delas estejam presentes no *corpus* de treinamento.

Dessa forma, defendemos que o *Lexicon* e o *Constructicon* precisam funcionar juntos para que a continuidade entre o léxico e gramática (FILLMORE, 2008) seja mapeada de forma satisfatória, uma vez que consideramos na tarefa de reconhecimento não apenas os *frames* das construções, como também as relações com as famílias de *frames* dos itens lexicais das construções anotadas.

A perspectiva de tratar os fenômenos semânticos via frames e formalizar o significado das sentenças juntamente com seus aspectos sintáticos fornece à modelagem de construções apresentada nesse trabalho um alinhamento entre o modelo de descrição linguística da Gramática das Construções de Berkeley e a implementação computacional da FN-Br. Entretanto, ambos os recursos estão ainda em desenvolvimento, por isso, não abrangem muitas ocorrências construcionais e lexicais do Português do Brasil. Apesar da limitação de extensão, a FN-Br tem buscado a ampliação do seu banco de dados, investindo na catalogação de construções nucleares e periféricas do PB.

7 CONCLUSÃO

Este trabalho modelou 24 construções de estrutura argumental de período simples do Português Brasileiro no *Constructicon* da FrameNet Brasil, com base nos pressupostos teóricos da Gramática de Construções de *Berkeley* e da Semântica de *Frames*. Com base nessa modelagem, procedeu à anotação de instâncias dessas construções no *corpus UD_Portuguese-Bosque*, com o intuito de submeter essas anotações a uma tarefa de identificação de construções por modelos de IA. Diante desse cenário, é necessário que façamos algumas considerações no que tange às implicações e contribuições desta tese para a descrição do PB e para a Linguística Computacional.

O primeiro ponto que trazemos para discussão nesta seção diz respeito ao foco deste trabalho na interação entre a Linguística e a Computação a fim de propor a modelagem de construções através do *Constructicon* do PB. Dessa forma, essa modelagem, que se encaixa dentro da fase representacional (DIAS DA SILVA, 1996), buscou representar formalmente as construções Sujeito Predicado, Transitiva Direta Ativa, Transitiva Indireta, Transitiva Oblíqua, Bitransitiva, Intransitiva, Ergativa, Resultativa, Objeto Interdito, Argumento Cindido Artefato, Argumento Cindido Parte do Corpo, Argumento Cindido Posse Alienável, Argumento Cindido Entidade Atributo, Predicativa Nominal Atributiva, Predicativa Nominal Estativa, Predicativa Locativa, Verbo Nominal Objetiva, Mudança de Estado, Sem Sujeito Intransitiva, Existencial Apresentacional e Tempo Decorrido, tornando-as computacionalmente tratáveis.

Como a modelagem e os resultados obtidos apontam para a importância da interação entre os recursos disponíveis na FN-Br, a unificação entre *Lexicon* e *Constructicon* permitiu que os aspectos sintáticos e semânticos das construções fossem descritos de forma equânime. Os aspectos semânticos das construções foram tratados a partir da perspectiva da semântica de *frames* e a formalização do significado das sentenças juntamente com seus aspectos sintáticos proporcionou à modelagem de construções apresentada nesse trabalho um alinhamento entre o modelo de descrição linguística da Gramática das Construções de *Berkeley* e a implementação computacional da FN-Br.

Como a nossa modelagem teve foco nas construções de estrutura argumental de período simples, muitas estruturas não foram abarcadas nesse estudo. Se, por um lado, reconhecemos nesse fato uma limitação, por outro, percebemos também que conseguimos tratar computacionalmente tanto daquelas construções consideradas mais nucleares – como a Transitiva Direta Ativa, Transitiva Oblíqua e a Predicativa Nominal Atributiva –

quanto de construções ditas mais periféricas como a Resultativa, a Mudança de Estado e Objeto Interdito. Mesmo que algumas construções modeladas não apresentassem ocorrências no *corpus*, quando propusemos o trabalho de modelagem dessas estruturas, nosso objetivo não era apenas testar essa modelagem para a identificação de construções, mas também propor um tratamento computacional a partir dos recursos existentes FN-Br que pode ser útil em outras tarefas de NLU futuras, além de constituir um valioso recurso para pesquisadores do PB. Para além das 24 construções de estrutura argumental em si, a modelagem incluiu outras 22 construções, dentre as que representam tipos sintagmáticos e classes de palavras, por exemplo.

O *Constructicon* abarca a descrição de construções em termos de características gramaticais e semânticas, realizando descrições que consideram os principais aspectos de cada construção. A formalização desse modelo de gramática que adotamos ocorre através do processo de unificação. O modelo de língua usado na representação das construções realiza uma combinação entre as construções e seus signos filhos resultando assim em uma Matriz de Valor e Atributo (AVM), a qual é constituída dos valores e atributos das AVMs que se uniram. Através de um conjunto de traços, essas AVMs explicitam as características de cada elemento, bem como as possibilidades (ou restrições) de combinações de dois ou mais elementos em uma construção.

Um aspecto significativo para a modelagem diz respeito às restrições construcionais que, nesse trabalho, englobaram desde os aspectos semânticos aos sintáticos dos signos filhos – ou CEs. Assim, as restrições sintáticas foram tratadas em termo dos CEs que são representados em sintagmas e categorias gramaticais como uma forma de mostrar as generalizações gramaticais de cada construção. Para que fosse possível possibilitar o mesmo tratamento tanto para a sintaxe quanto a semântica das construções, os CEs também foram unificados a um determinado componente semântico, ou seja, um EF do frame evocado pela construção. Dessa forma, os resultados da modelagem nos permitiram obter uma rede de construções por meio das seguintes relações: Herança – entre construções –, Evocação – entre construções e frames e Restrições, necessárias para apontar características mais específicas de cada construção e assim modelar as construções aqui estudadas.

A modelagem de 24 construções todas herdeiras em alguma medida da construção Cláusula resultou em uma rede em que foi possível perceber como as construções gramaticais estão relacionadas entre si através de diferentes aspectos. Nesse contexto, foi descrito de que forma as propriedades das construções que foram aqui expostas em

diálogo com diversas descrições linguísticas estão relacionadas com outras construções que contribuem para sua definição, seja através das relações de herança, ou de restrições, que estão relacionadas com a composição de cada construção.

A rede construcional evidenciou que as construções em estudo apresentaram uma relação de herança entre a construção mais genérica Cláusula, que, por sua vez, é herdeira da Construção X_Núcleo e Predicado, sendo a primeira relacionada via Herança as construções Sujeito_Predicado e a segunda relacionada às construções Sem Sujeito.

Outra contribuição da modelagem foi a realização de um mapeamento que permitiu que diferentes construções que apresentam uma estrutura semelhante fossem discriminadas computacionalmente conforme suas características particulares, pelo mapeamento das restrições de cada construção. Temos, como exemplo, as construções Transitiva Direta Ativa e Argumento Cindido, que apresentam uma mesma estrutura sintática, porém são distintas semanticamente tendo a primeira um caráter causativo e a segunda um caráter inacusativo, que, em certo ponto, se assemelha com a construção Ergativa.

Podemos destacar, ainda, as construções Ergativa, Intransitiva e Objeto Interdito, que, apesar de serem estruturalmente monorgumentais, apresentam significados diversos. Enquanto na construção Ergativa temos uma entidade sendo afetada por um evento, a construção Intransitiva diz respeito a um ato intencional praticado pelo sujeito e a Objeto Interdito traz um verbo prototipicamente transitivo cujo objeto se encontra omitido devido ao seu caráter pejorativo.

Ao observarmos as construções de argumento preposicionado, como é o caso da Transitiva Indireta, Transitiva Oblíqua e Bitransitiva, percebemos como a construção Transitiva Oblíqua se comporta diferentemente da Transitiva Indireta, já que comumente o objeto oblíquo é tratado como objeto indireto. Nas construções Transitiva Indireta e Bitransitiva, o *frame* Transferir, evocado por ambas, assume perfilamentos diferentes de acordo com a construção. Se, por um lado, a construção Transitiva Indireta deixa mais em evidência o Receptor, por outro, a construção Bitransitiva perfila os dois EFs do *frame* em questão, Receptor e Tema.

Com a modelagem das construções Predicativas foi possível tratar da distinção da natureza dos predicativos do sujeito e dos verbos chamados cópulas. Na Predicativa Nominal Estativa consideramos através das restrições que o *slot* da cópula não pode ser preenchido pelo verbo *ser* devido a sua noção de permanência, o que não se observa na construção Predicativa Nominal Atributiva. Já na Predicativa Locativa o EC Predicado

tem que ser constituído ou por Sintagma Adverbial ou um Sintagma Preposicionado diferentemente das Predicativas nominais em que esse *slot* tem que ser preenchido por um Sintagma Nominal ou Sintagma Adjetival.

Nas construções Verbo Nominais, mapeamos como os considerados predicados verbos nominais podem ser tratados através de uma perspectiva construcional, uma vez que não tratamos essas construções sob o viés da predicação secundária, mas propusemos que os novos sentidos que essas estruturas assumem são uma contribuição da construção e não só dos elementos lexicais, o mesmo acontecendo para as construções Resultativas.

Por fim, as construções clausais Sem sujeito foram modeladas para proporcionar um tratamento computacional a esse tipo de estrutura existente no PB. A construção Sem Sujeito Intransitiva foi considerada como responsável por licenciar construtos que indicam fenômenos da natureza, já as construções de Tempo Decorrido foram consideradas monoargumentais marcadas pela presença de um verbo impessoal acrescido de uma complementação que, na Existencial Apresentacional, tem um caráter apresentativo e, na Tempo Decorrido, locativo ou temporal.

Uma segunda contribuição do trabalho diz respeito à submissão da modelagem realizada a uma tarefa de reconhecimento automático de construções em PB. Assim, a fase implementacional consistiu em uma tarefa de identificação construcional. Em um primeiro momento foram realizadas anotações, no corpus *UD_Portuguese-Bosque*, das construções que foram modeladas no *Constructicon*. As construções modeladas e anotadas no *corpus* foram submetidas a distintos modelos de Inteligência Artificial para serem classificadas conforme a modelagem.

Os dados indicaram, em uma primeira análise, que o desempenho do modelo de IA melhora quando são levados em consideração os traços relativos ao polo formal das construções. Já quando informações relativas ao polo semântico são usadas, o desempenho dos modelos diminui. Entretanto, uma análise de desempenho para cada 749 nova análise conseguimos perceber como o processo de modelagem semântica em alguns casos traz ganho de resultado para a identificação de construções, especialmente quando essas construções não são tão frequentes no *corpus* de anotação. Foi possível identificar como os modelos que traziam novas informações conseguiram identificar ECs até então não identificados e conseguiram melhorar a identificação de algumas construções que não estavam sendo reconhecidas. Isso sugere que, quando temos construções pouco frequentes no *corpus* de treinamento, as informações adicionadas pelos modelos conseguem se apresentar como dados significativos para tarefas de identificação. Sendo

assim a modelagem para construções menos frequentes, ou em um cenário de escassez de dados, pode contribuir para esse tipo de tarefa.

Portanto, esta tese, a partir das análises empreendidas, propõe os seguintes avanços: (i) a partir das discussões teórico-metodológicas empreendidas neste trabalho, propôs-se a modelagem de 24 construções para um empreendimento identificação de construções no PB; (ii) modeladas as construções, apresentaram-se os resultados de uma tarefa de identificação de construções que utiliza modelos de IA. Dadas as limitações do *corpus* e do *Constructicon*, deixa-se a abertura para investigações futuras, as quais, partindo do alinhamento entre os recursos da FN-Br e algoritmos de IA, visem desenvolver outras tarefas de identificação de construções a partir de *corpora* de treinamentos maiores anotados com as categorias criadas neste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALLEN, J. **Natural Language Understanding**. New York: Pearson, 1995.
- ALMEIDA, V. G. **Identificação Automática de Construções de Estrutura Argumental: um experimento a partir da modelagem linguístico-computacional das construções Transitiva Direta Ativa, Ergativa e de Argumento Cindido**. Dissertação de Mestrado em Linguística. Universidade Federal de Juiz de Fora, Juiz de Fora, 2016.
- BÄCKSTRÖM, Linn et al. Automatic identification of construction candidates for a Swedish constructicon. **Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA**, 2013. NEALT Proceedings Series 19 / Linköping Electronic Conference Proceedings 88: 2–11, 2013.
- BAILEY, David R. **When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs**. Ph.D. thesis, University of California at Berkeley, 1997.
- BARBOSA, J. W. C. Ter estado resultante não é ter construção resultativa: predicados secundários pseudo-resultativos e orações adjuntas de resultado no português brasileiro. **DELTA: Documentação e Estudos em Linguística Teórica e Aplicada**, [S. l.], v. 34, n. 2, 2018.
- BENDER, Emily M. e Alexander KOLLER. **L Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data**. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 5185–5198, Online. Association for Computational Linguistics, 2020
- BERGEN, Benjamin e CHANG, Nancy. **Embodied Construction Grammar**. Oxford Handbook of Construction Grammar, 2004.
- BERGEN, Benjamin e WHEELER, Kathryn. **Sentence understanding engages motor processes**. Cognitive Science – COGSCI, 2005.
- BRONZATO, L. H. O enquadre gramatical da interdição ou “a bom entendedor meia palavra basta”. In: Miranda, N. S. & Salomão, M. M. M. (orgs.). **Construções do Português do Brasil: da gramática ao discurso**. Belo Horizonte: UFMG, p. 76- 97, 2009.
- BRUNYÉ, Tad, DITMAN, Tali, MAHONEY, Caroline e TAYLOR, Holly. **Better you than I: Perspectives and emotion simulation during narrative comprehension**. Journal of Cognitive Psychology. 23. 10.1080/20445911.2011.559160, 2011.

- BRYANT, John. **Best-Fit Constructional Analysis**, PhD. dissertation, UC Berkeley, 2008.
- BOAS, Hans C. **A Constructional Approach to Resultatives**. Stanford: CSLI Publications, 2003.
- BORGES, V. E, WENDT, Maximilian, HENNIG, Felix, e KÖHN, Arne. **HDT-UD: A very large Universal Dependencies Treebank for German**. In Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019), pages 46–57, Paris, France. Association for Computational Linguistics, 2019.
- CASTILHO, A. T. **Nova Gramática do Português Brasileiro**. São Paulo: Editora Contexto, 2010.
- CHOMSKY, N. **Aspects of theory of syntax**. Cambridge, MA: MIT Press. 1965.
- CROFT, W. **Radical construction grammar: syntactic theory in typological perspective**. Oxford: Oxford University Press, 2001.
- COSTA, A. D. & TORRENT, T.T. **A Modelagem Computacional do Domínio dos Esportes na FrameNet Brasil**. In: Proceedings of Symposium in Information and Human Language Technology. Ubelândia, MG. Brazil. Sociedade Brasileira de Computação, p. 201-208, 2017.
- CUNHA, M. A. F. D. **Motivações semântico-pragmáticas para a ordenação dos argumentos na construção ditransitiva**. Revista de Estudos da Linguagem, v. 25, p. 555-584, 2017.
- CUNHA, C. F. & CINTRA, L. **Nova Gramática do Português Contemporâneo**. Rio de Janeiro: Nova Fronteira, 2013.
- DANCYGIER, Barbara e Sweetser, Eve. **Mental Spaces in Grammar: Conditional Constructions**. Cambridge, England: Cambridge University Press, 2005.
- DEVLIN, Jacob, CHANG, Ming-Wei, LEE, Kenton, e TOUTANOVA, Kristina. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics, 2019.
- DIAS-DA-SILVA, B. C. **A face tecnológica dos estudos da linguagem: O processamento automático das línguas naturais**. Araraquara, 1996. Tese de Doutorado em Letras - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 1996.

- DIAS-DA-SILVA, B. C. **O estudo Linguístico-Computacional da Linguagem**. In: Letras de Hoje, v. 41, n. 2, p. 103–138, 2006.
- DINIZ, Alexandre & ALMEIDA, Vânia, LAGE, Ludmila, BARBOSA, Gustavo, MARÇÃO, Natália, PAIVA, Vanessa, MATOS, Ely e TORRENT, Tiago. **Representação computacional das construções de sujeito-predicado do português do Brasil**. Revista Linguística. 14. 149. 10.31513/linguistica.v14n1a15285, 2018
- DUNN, Jonathan. **Computational Learning of Construction Grammars**. Language and Cognition 9 (2): 254–292. 10.1017/langcog.2016.7, 2017.
- FAUCONNIER, G. **Mappings in language and thought**. Cambridge: Cambridge University Press, 1997.
- FAUCONNIER, G. **Mental spaces**. Cambridge: Cambridge University Press, 1994.
- FAUCONNIER, G.; SWEETSER, E. (Org.). **Spaces, words and grammar**. Chicago: Chicago University Press, 1996.
- FAUCONNIER, G.; TURNER, M. **The way we think**. New York: Basic Books, 2002
- FELDMAN J. A. **From Molecule to Metaphor: A Neural Theory of Language**. Cambridge: The MIT Press, 2006.
- FELDMAN, J. A. **On the Evolution of Subjective Experience**, 2020.
- DODGE, E. K.. **Constructional and Conceptual Composition**. PhD. diss, Department of Linguistics, University of California, Berkeley, 2010.
- FERREIRA, M. A Construção de Ação Rotineira no Português do Brasil. In__MIRANDA, N. S.; SALOMÃO, M. M. M. **Construções do Português do Brasil: da gramática ao discurso**. Belo Horizonte: UFMG, 2009.
- FILLMORE, C. J. **The case for case**. In: BACH, E. & HARMS, R. (eds) *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston, 1968.
- _____. **The case for case reopened**. In: COLE, P.; SADDOCK, J. (Org.). *Grammatical relations*. New york: Academic, 1977a.
- _____. **Topics in lexical semantics**. In: COLE, R. (Org.). *Current issues in linguistic theory*. Bloomington: Indiana University Press, 1977b.
- _____. **Innocence: a second idealization for linguistics**. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: BLS, v. 5, p. 63-76, 1979.
- _____. **Frame semantics**. In: Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*. Seoul: Hanshin, p.111-138, 1982.
- _____. **Frames and the Semantics of Understanding**. *Quaderni di semantica*, 6 (2), p. 222- 254, 1985.

_____. **Border conflicts: FrameNet meets construction grammar**. EURALEX,13, 2008, Barcelona. Anais. Barcelona: Universitat Barcelona Fabra, 2008.

_____. Berkeley Construction Grammar. In: HOFFMANN, T.; TROUSDALE, G. 77 **Oxford Handbook of Construction Grammar** (Eds.). Oxford University Press, 2013.

FILLMORE, Charles; KAY, Paul; O'CONNOR, Mary Catherine. **Regularity and idiomaticity in grammatical constructions: the case of 'let alone'**. *Language* 64, p. 501-38, 1988.

FILLMORE, Charles, JOHNSON, Christian. R. e PETRUCK, Miriam R. L. **Background to FrameNet**. *International Journal of Lexicography*, 16 (3): 235- 250, 2003.

FILLMORE, C. J., LEE-GOLDMAN, R. e ROMIEUX, R. The FrameNet Constructicon. In: BOAS, H. C. and SAG, I. A. (Eds.). **Sign-based Construction Grammar**. Stanford: Center for the Study of Language and Information, 2012. p.309-372.

FOLTRAN, M. J. **As construções de predicação secundária no português do Brasil: aspectos sintáticos e semânticos**. Tese de Doutorado em Linguística. Universidade Federal do Paraná, Curitiba, 1999.

FRIED, M. & ÖSTMAN, J-O. Construction Grammar: A thumbnail sketch. In: _____. (orgs.). **Construction Grammar in a Cross-Linguistic Perspective**. Amsterdam: John Benjamins, p. 11–86, 2004.

GOLDBERG, A. E. **Constructions: A Construction Grammar Approach to Argument Structure**. Chicago: Chicago University Press, 1995.

_____. **Constructions at Work: The nature of generalization in language**. Oxford: Oxford University Press, 2006.

GOLDBERG, Adele; JACKENDOFF, Ray. **The resultative as a family of constructions**. *Language* 80, p. 532-568, 2004.

Honnibal, Montani spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing, 2017.

KAY, P & FILLMORE, C. J. **Grammatical constructions and linguistic generalizations: The 'What's X doing Y?'** Construction. *Language*, 75 (1), p.1–33,1999.

LAGE, L. M. **Frames e Construções: A Implementação do Constructicon na Framenet Brasil**. Dissertação de Mestrado apresentada ao programa de Pós-Graduação em Linguística da Faculdade de Letras da Universidade Federal de Juiz de Fora, 2013.

LAGE. L. M. **Modelagem Linguístico-Computacional das Relações entre**

Construções e Frames no Constructicon da FrameNet Brasil. Tese de Doutorado em Linguística. Universidade Federal de Juiz de Fora, Juiz de Fora, 2018.

LAKOFF, G. **The contemporary theory of metaphor.** In: ORTHONY, A. (Org.). *Metaphor and thought.* 2.ed. Cambridge: Cambridge University Press, 1983.

_____. **Women, fire and dangerous things.** Chicago: Chicago University Press, 1987.

LAKOFF, G. & JOHNSON, M. **Metaphors we live by.** Chicago: Chicago University Press, 1980.

_____. **Philosophy in the flesh.** New York: Basic Books, 1999.

_____. **Why cognitive linguistics requires embodied realism.** *Cognitive Linguistics*, [S. l.], v. 13, n. 3, p. 245-263, 2002.

LANGACKER, R. **Foundations of cognitive grammar.** v. 1. Stanford: Stanford University Press, 1987.

_____. **Foundations of cognitive grammar.** v. 2. Stanford: Stanford University Press, 1991.

LAVIOLA, Adrieli Bonjour da Silva Darowisch, **Constructicografia multilíngue em ação: diretrizes linguístico-computacionais para o alinhamento de constructicons.** Tese de Doutorado em Linguística. Universidade Federal de Juiz de Fora, Juiz de Fora, 2019.

LYNGFELT, B., BORIN, L. FORSBERG, M., PRENTICE, J., RYDSTEDT, R., SKÖLDBERG, E. e TINGSSELL, S. **Adding a Constructicon to the Swedish resource network of Språkbanken.** *Proceedings of KONVENS 2012 (LexSem 2012 workshop)*, Wien, pp. 452-461, 2012.

MADABUSHI, Harish Tayyar, ROMAIN, Laurence, DIVJAK, Dagmar e MILIN, Petar. **CxGBERT: BERT meets Construction Grammar.** In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4020–4032, Barcelona, Spain (Online). International Committee on Computational Linguistics, 2020.

MARQUES, Tânia e BEULS, Katrien. **Evaluation Strategies for Computational Construction Grammars.** In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1137–1146, Osaka, Japan. The COLING 2016 Organizing Committee, 2016.

MASSON, Michael; BUB, Daniel; WARREN, Christopher. **Kicking calculators: Contribution of embodied representations to sentence comprehension.** In: **Journal of Memory and Language**, 2008, n.59, p. 256-265.

- MATOS, E. E. da S. **LUDI: um framework para desambiguação lexical com base no enriquecimento da Semântica de Frames**. Tese de Doutorado em Linguística. Universidade Federal de Juiz de Fora, Juiz de Fora, 2014.
- MCRAE, Ken; SPIVEY-KNOWLTON, Michael J.; TANENHAUS, Michael K. **Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension**. *Journal of Memory and Language*, v. 38, n. 3, p. 283-312, 1998.
- MCSHANE, M. Natural Language Understanding (NLU, not NLP) in **Cognitive Systems**. *AI Magazine*, [S. l.], v. 38, n. 4, p. 43-56, 2017. DOI: 10.1609/aimag.v38i4.2745., 2017.
- NARAYANAN, Srini. **KARMA: Knowledge based active representation for metaphor and aspect**. Ph.D. Dissertation. Computer Science Division. University of California, Berkeley, 1997.
- NEWMAYER, Frederick J., **The politics of linguistics**. Chicago: The University of Chicago Press, Pp. viii 171. *Journal of Linguistics*, 24(1), 263-264. doi:10.1017/S0022226700011774, 1986.
- PERINI, M. A. **Gramática do Português Brasileiro**. São Paulo: Parábola, 2010.
- PUSTEJOVSKY, James. The Generative Lexicon. **Computational Linguistics**, 17(4):409–441, 1991
- PUSTEJOVSKY, J. e STUBBS, A. **Natural Language Annotation for Machine Learning: A guide to corpus building for applications**. Sebastopol, CA: O'Reilly, 2012.
- RAJPURKAR, Pranav, JIA, Robin, e LIANG, Percy. **Know What You Don't Know: Unanswerable Questions for SQuAD**. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 784–789, Melbourne, Australia. Association for Computational Linguistics, 2018.
- RAPPAPORT Hovav, Malka e LEVIN Beth. Building verb meanings. In Miriam Butt & Wilhelm Geuder (eds.), **The projection of arguments: Lexical and syntactic constraints**, 97–134, 1998.
- RAPOSO, E. B. P.; BACELAR DO NASCIMENTO, M. F.; MOTA, M. A. C. da; SEGURA, L.; MENDES, A. (Coord.). **Gramática do Português**. Lisboa: Fundação Calouste Gulbenkian. XLVII+XXIII+2409 p., 2013.
- ROCHA LIMA, Carlos Henrique da. **Gramática normativa da língua portuguesa**. 49.ed. Rio de Janeiro: José Olympio, 2007.

ROGERS, Anna, KOVALEVA, Olga e RUMSHISKY, Anna. **A primer in bertology: What we know about how BERT works**. CoRR, abs/2002.12327, 2020.

SALOMÃO, Maria Margarida Martins. 3) A questão da construção do sentido e a revisão da agenda dos estudos da linguagem. **Veredas-Revista de Estudos Linguísticos**, v. 3, n. 1, 1999.

SALOMÃO, Maria Margarida Martins. Lanterna na proa: sobre a tradição recente nos estudos da linguística. **Gragoatá**, v. 12, n. 23, 2007.

SALOMÃO, Maria Margarida Martins. FrameNet Brasil: um trabalho em progresso. **Calidoscópico**, 7(3), p. 171- 182, 2009a.

SALOMÃO, Maria Margarida Martins Tudo certo como dois e dois são cinco: Todas as construções de uma língua. In: MIRANDA, N. S. & SALOMÃO, M. M. M. (Org.). **Construções do Português do Brasil: da gramática ao discurso**. Belo Horizonte: UFMG, 2009b.

SALOMÃO, M.M.M.; SAMPAIO, T.F.; TORRENT, T.T. A linguística cognitiva encontra a linguística computacional: notícias do projeto FrameNet Brasil. **Cadernos Estudos Linguísticos**. Universidade Estadual de Campinas. Campinas. nº 55. p. 7-34, jan/jun, 2013.

SAMPAIO, T. F. **A Família de Construções de Argumento Cindido no Português do Brasil**. Tese de Doutorado em Linguística. Universidade Federal de Juiz de Fora, Juiz de Fora, 2010.

SAUSSURE, F. **Curso de linguística geral**. São Paulo: Cultrix, 1972.

SILVA, A. B. L.; LAGE, L. M.; MARCAO, N. D.; TAVARES, T. S.; ALMEIDA, V. G.; MATOS, E. E. & TORRENT, T. T. In: The AAI 2017 **Spring Symposium on Computational Construction Grammar and Natural Language Understanding** 79 Technical Report SS-17-02. Palo Alto, CA: AAI Publications, v.17, p.193-196, 2017.

STANFIELD, Robert A.; ZWAAN, Rolf A. The effect of implied orientation derived from verbal context on picture recognition. **Psychological science**, v. 12, n. 2, p. 153-156, 2001.

STEELS, Luc. Modeling the formation of language in embodied agents: Methods and open challenges. In: **Evolution of Communication and Language in Embodied Agents**. Springer, Berlin, Heidelberg, 2010. p. 223-233.

STEELS, Luc (Ed.). **Design patterns in fluid construction grammar**. John Benjamins Publishing, 2011.

- STEELES, Luc. Fluid construction grammar. In: **The Oxford handbook of construction grammar**. 2013.
- STEELES, Luc. Basics of fluid construction grammar. **Constructions and frames**, v. 9, n. 2, p. 178-225, 2017.
- TALMY, L. **Figure and ground in complex sentences**. In: GREENBERG, J. et al. **Universals of human language**. Stanford: Stanford University Press, 1978.
- TALMY, L. **How languages structure space**. In: PICK, H.; ACREDOLO, L. (Org.). **Spatial Orientation**. New York: Plenum Press, 1983.
- TORRENT, T. T. & ELLSWORTH, M. Behind the Labels: criteria for defining analytical categories in FrameNet Brasil. **Veredas**, 17 (1), p. 44-65, 2013.
- TORRENT, T. T. ; MATOS, E. E. ; BELCAVELLO, F. ; VIRIDIANO, M. ; GAMONAL, M. A. ; DINIZ DA COSTA, A. ; MARIM, M. C. . Representing Context in FrameNet: a multidimensional, multimodal approach. **Frontiers in Psychology**, v. 13, p. 838441, 2022.
- TSAO, Nai-Lung e WIBLE, David. A Method for Unsupervised Broad-Coverage Lexical Error Detection and Correction. In **Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications**, pages 51–54, Boulder, Colorado. Association for Computational Linguistics, 2009.
- WANG, Alex et al. GLUE: A multi-task benchmark and analysis platform for natural language understanding. **arXiv preprint arXiv:1804.07461**, 2018.
- ZIEM, Alexander; FLICK, Johanna; SANDKÜHLER, Phillip. The German Constructicon Project: Framework, methodology, resources. **Lexicographica**, v. 35, n. 2019, p. 15-40, 2019.
- ZWAAN, Rolf A.; STANFIELD, Robert A.; YAXLEY, Richard H. Language comprehenders mentally represent the shapes of objects. **Psychological science**, v. 13, n. 2, p. 168-171, 2002.